DOCUMENT RESUME

ED 192 165 CE 027 160

AUTHOR Harlan, Anne: And Others
TITLE The Assessment of Occupational Competence. 1. Competence Assessment in Personnel Selection: Current Practices and Trends.
INSTITUTION McBer and Co., Boston Mass.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Feb 80
CONTRACT 400-78-0028
NOTE 191p.; For related documents see CE 027 159-165.

EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS Adults; *Competence; Employment Qualifications; Evaluation Methods; Job Performance; *Job Skills; Minimum Competency Testing; *Occupational Tests; *Personnel Evaluation; *Personnel Selection; Predictive Measurement; Vocational Aptitude
IDENTIFIERS *Occupational Competence Assessment

ABSTRACT

One of seven sections of a report that examines the assessment of occupational competence, this chapter presents competence assessment as it is defined in practice by selection techniques currently used by employers. The chapter begins with a discussion of the major techniques used in employee selection, focusing on the competencies employers detect and their measurement properties. The second major portion focuses on the ways in which selection techniques are used as components of selection systems within a variety of organizations. Among the questions addressed here are the following: Do employers actually select on the basis of the competencies they desire in job applicants? How do selection practices vary according to the type of job and level of skill required at entry? What biases enter into selection decisions as a function of the procedures used? To what extent do employers verify the importance of certain competencies to job performance or the effectiveness of their selection procedures? The chapter concludes with a discussion of how organizational realities and other factors affect selection practices and what educators can learn from current trends in these practices. (Other sections of the report are available separately--see note. The first is an overview; the last is a synthesis of issues.) (LRA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED192165

THE ASSESSMENT OF OCCUPATIONAL COMPETENCE

1. COMPETENCE ASSESSMENT IN PERSONNEL SELECTION: CURRENT PRACTICES AND TRENDS

Anne Harlan
George O. Klemp
Mary Lou Schaalman

Prepared for
The National Institute of Education
Contract NIE 400-78-0028

February, 1980

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

McBER 137 NEWBURY STREET · BOSTON · MASSACHUSETTS 02116

CE027160

2

DEC 4 1980

TABLE OF CONTENTS

		page
I.	Introduction and Overview	1.1
II.	Review of Employee Selection Techniques	1.10
	Selection Interviews	1.12
	Psychometric Tests	1.27
	Other Selection Devices	1.53
	Differential Validity and Test Bias	1.78
	Comparison of Selection Devices	1.91
III.	The Practice of Personnel Selection	1.99
	The Scope of the Study	1.100
	Trends in Contemporary Practice	1.106
	Analysis of Trends and Practices	1.114
IV.	Conclusions	1.133
	Discrepancy Between Actual and Ideal Practice	1.133
	Factors Underlying the Discrepancy	1.136
	Implications for Educators and Policymakers	1.148
References		1.165

I. INTRODUCTION

In a competitive economy where more than one applicant may be found for most jobs or career opportunities that are made available, there exists the need for mechanisms to select the most qualified applicants out of the rest of the competition. From an employer's point of view, a selection procedure should be able to identify people who are right for the job as well as to ensure that selected individuals will choose to remain in the job long enough to justify the employer's investment in training. The degree of success in meeting these aims varies greatly with the selection procedure used and with the requirements of the job to be filled.

Nearly all the selection procedures that are currently used were in use more than 30 years ago. Today, many of these procedures evidence a greater degree of sophistication in their design and application than that which was tolerated in a more naive age. Thirty years ago, subjective judgments on the part of the interviewing employer, a bachelor's degree from a "good school", and test scores on a variety of psychometric measures of general ability all had a life of their own; the relationship between these factors and the merit of applicants for available jobs remained, with few exceptions, largely unquestioned outside of academic circles. Indeed, very few studies

of the cost-effectiveness of these devices were undertaken because of their appearance of validity and because the costs of such studies in both dollars and inconvenience outweighed the costs of administering the selection systems themselves.

Title VII of the Civil Rights Act of 1964 ushered in a new era and a new consciousness with regard to employee selection. Title VII is the portion of the Civil Rights Act that deals with discrimination in employment. It is especially relevant to employment selection processes because of Section 703(h), known as the Tower Amendment, which states that it is not "an unlawful employment practice for an employer to give and to act upon the results of any professionally developed ability test provided that such test, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex, or national origin." This section was added to the Act as a result of concern in Congress that companies would be prevented by the courts from using employment tests which disqualified members of protected classes (i.e., those identifiable by race, color, religion, sex, or national origin). The phrase "professionally developed test" became the impetus for the 1970 version of the EEOC selection guidelines. The guidelines served to define what a test is, what constitutes discriminatory use of tests, and what standards of validity should be used to judge whether a test has been developed carefully, fairly, and rigorously. These guidelines, in turn, became the touchstone, implicitly and explicitly, for professional practice and legal decisions.

Employers were now compelled to demonstrate statistically or rationally the relationship between selection procedures and jobs, to develop procedures which met proposed standards of validity, and to carefully document the fairness of these procedures to previously protected groups. Many organizations began to shy away from testing either out of an inability to deal with the cost and rigor of validation or out of the realization that their procedures did not measure up to the guidelines. In 1972, when an amendment to Title VII gave the EEOC enforcement power, the use of certain selection procedures abated even more swiftly. Ironically, psychometric tests of ability and personality, those selection devices for which statistical data were most readily available, were the first devices to be abandoned, since they were also the first to show discriminatory impact. On the other hand, other devices such as the selection interview, while generally much lower in reliability and validity, have increased in use, since it is more difficult to document the adverse impact of such devices.

At the same time, the last ten years have etched a growing concern in the minds of employers regarding the efficiency and productivity of the work force. Fierce competition from abroad, where labor was less costly and where technology had caught up with the domestic standard in many areas, contributed initially to this concern. In more recent years, the slowing of economic growth due to currency inflation, the increasing cost of energy, and a decline in the availability of resources

have placed additional emphasis upon productivity as the key issue for the American economy. The recent consumer movement has created its own pressures for assuring the value and quality of goods and services, and the passage of the Consumer Protection Act in 1974 placed additional costs on many organizations which, in the short run, exacerbated the problem of maintaining costs for goods and services delivered. These sources of pressure have placed an additional burden on personnel selection systems to be cost-effective means of selecting competent and productive individuals, and to screen out those whose lower productivity and shorter tenure make them poor hiring risks.

The employer of today is faced with a critical dilemma. Given the state of the art of psychological measurement, is it possible to use a selection procedure that enhances the ability of an organization to employ individuals who are both competent and highly productive, while not excluding disproportionately individuals on the basis of characteristics such as age, race, or sex which are, in theory, not correlated with job performance? Have employers pursued advancing the art and the practice of selection to make selection both more fair and more predictive of effective performance? Or have they turned away from selection systems as devices for ensuring competence and toward on-the-job training and development?

This chapter is devoted to identifying the state of the practice in competence assessment in employer selection. It

proceeds with a review of selection techniques in current use, with particular emphasis on variations, validities, and problems associated with each. The section that follows shifts focus to the selection practice in more than 200 jobs which represent the spectrum of career opportunity. The concluding section responds to three critical issues for the practice of personnel selection: the discrepancy between actual and ideal practice; some of the factors underlying the discrepancy; and implications of the state of the practice for both educators and employers. These issues are introduced in greater detail below so that they might serve as a conceptual overview to the discussions which follow.

Discrepancy Between Actual and Ideal Practice

Competence assessment for selection can be viewed from two perspectives, that of the personnel researcher, and that of the personnel practitioner in a functioning organization. The ideals of the former perspective might be unlikely to converge with the realities of the latter. In fact, we expected to find considerable discrepancy between the ideal and the actual practice. This disparity is all the more fruitful for our research because it suggests that the selection practices in use today can be more controlled, reliable and valid, and that the existing research on selection practices can be set as a minimum standard for needed improvement.

While the research literature can be used as a base for evaluating the state of the practice, what can serve as a standard for evaluating competence assessment research itself? Although there is no absolute standard for assessing the quality of research, over the years several researchers have written prescriptive documents highlighting the parameters of predictors, measures, samples and criteria which will lead to optimal selection reliability and validity. These prescriptions remain appropriate for evaluating the state of the practice today. In addition, research that neglects these longstanding caveats in the designing of studies is open to criticism. Comparisons between research and practice, with regard to accepted standards of excellence, therefore, will be implicitly and explicitly stated throughout this chapter.

Factors Underlying the Discrepancy

As already noted, competing pressures on employers and the varying efficiency of employee selection procedures are expected to have a telling effect on the practice of competence assessment. Other factors which will later be brought into the picture include the difficulty and practicality of conducting studies on the validity of selection procedures within organizations and characteristics of the organizations themselves that contribute to the likelihood that validity studies will be undertaken. There is always a period of delay between any process innovation and its application, and discussion will be

devoted to an explanation of factors contributing to delay in the use of state-of-the-art employee selection devices.

Indeed, some of the questions that bear on impediments to assimilating new assessment technology are relevant to implementing any new technology. (1) Is the organization ready for the innovation? Can it afford a dramatic change in present practice, and will top decision-makers in the organization support such a change? (2) Is the latest technology really much better than that which is already in use? Do suspicions of change for its own sake create resistance to innovation, or are the benefits of change seen as being too marginal to warrant it? (3) Do the costs of implementing the latest technology outweigh the benefits to be derived? Regarding employee selection, do the costs of validation of a new selection device reduce the likelihood that a well-documented long-used procedure which has wide organizational acceptance will be replaced? These issues flow logically from a discussion of the discrepancy between ideal and actual practices in employee selection, and will be dealt with on the basis of data gathered from the organizations surveyed in the analysis of job selection procedures.

Implications for Educators and Employers

As the present study was undertaken, it soon became evident that exceedingly few employers based their requirements for

employee selection on competency measurements. Whether or not an employer has been able to demonstrate a rational, valid relationship between job performance and specified knowledge, skills, abilities or other personal characteristics of the individual, the majority of employers do not attempt to measure these characteristics directly through an employee selection process. Rather, they rely on their ability to make accurate inferences about the presence or absence of job related competencies from data collected through devices such as the interview, personality inventory, or resume. Employers have rarely been in the practice of making explicit the competencies they seek, except insofar as they are defined by the selection procedures they choose to invoke. With the exception of certain skill tests and job simulations which provide close correspondence to the actual work to be performed, most other sources of information considered in the selection process are open to a great deal of interpretation in establishing a link between, for example, a test score or an event in the employee's background and a competency required for the job.

These observations have certain critical implications for both employers and educators. The employee selection procedures in current use are often the only indicators of the competencies required for jobs to which educators and job applicants have access. It is important, therefore, to examine what, if anything, selection procedures tell us about job requirements that is either true or misleading. Do employers

select applicants on the basis of competencies that are needed for satisfactory job performance, and if so, should educators better prepare students to survive selection assessment procedures? Or do employers make selection decisions on the basis of competencies that are largely irrelevant to job performance, and must educators therefore prepare students for both long-term career effectiveness and short-term desirability as a job applicant? The answers to these questions will enable us to examine the emerging role of the employing organization as an educator of adults, and to contrast that role with the one which is currently being played by the approach to education practiced by a number of contemporary "competency-based" educational institutions.

II. REVIEW OF EMPLOYEE SELECTION TECHNIQUES

This section is devoted to selection devices, or screening devices that employers apply to job applicants at entry into organizations. Promotion and performance appraisal systems, which are used as the bases for selection from within an organization, are not considered here, though many of the techniques described are applicable to this purpose. The interview and the psychometric test are the two most popular selection techniques, and separate sections will be devoted to discussions of each. In addition, six other devices--application blanks, resumes, recommendations, work samples and simulations, thought samples, and detectors of deception are currently in widespread use and a single section will be devoted to their description. Each of the sections will describe different forms of the devices, the range of competencies and other factors measured, the reliability and validity of the devices, and issues associated with their use.

Of particular importance in this review is the consideration of the predictive validity of a selection technique: to be competency-based, a selection device should measure qualities of the applicant that are related to effective job performance. Presently, employers consider four criteria in determining the validity of a selection technique, which vary in the degree to

which they can be regarded as indicative of employee competence. The reader should keep these criteria in mind as he or she considers the utility of a selection device for competency measurement.

1. "Hard" measures, or direct measure of employee performance such as sales volume, production error rate, and portfolio profitability, are to be preferred as criteria, where available, against which a selection system would be validated.

2. "Soft" measures, including supervisor and peer ratings of performance, are often collected in the absence of direct performance measures. Ratings are susceptible to reliability problems and may, in many cases, bear a low relationship with harder performance measures (Kane and Lawler, 1979).

3. Trainability, defined in practice as how well an employee performs in a training program offered by the employing institution, may be a useful criterion for the employer in keeping costly attrition low. However, whether trainability reflects job competence depends on the to which that the results achieved through training are statistically related to job performance.

4. Attrition, or employee turnover, is a valuable criterion to the employer who desires to select employees who will remain on the job. This criterion is unrelated to competence in the job.

Selection Interviews

The employment interview has proved to be the most widely used technique for employee selection. A survey by Scott, et al. (1961) found that 98.4% of the 852 organizations sampled used the interview in their selection process. Of the companies surveyed 93.9% responded that an applicant would never be hired without first conducting a personal interview. There is no evidence to suggest that widespread use of the selection interview has abated. Campbell, et al., (1970) found that of 146 organizations surveyed, all placed "great importance" on the interview as a selection technique, with only one company attempting to decrease the emphasis placed on it. Perhaps because of such widespread use, the selection interview has evolved into a plethora of forms and styles of which only the most popular will be discussed here.

Variety in Format

Structure

Three primary types of interview structure have been noted in the selection interview: the structured, semi-structured ("guided" or "patterned"), and the unstructured ("laissez-faire") interview. These three approaches differ in the extent to which they rely on a standardized set of questions in conducting the interview.

1. The most prevalent form of the selection interview according to Hakel (1977), is the semi-structured or patterned approach in which the interviewer covers certain broad areas of questioning, such as education, work experience, and past accomplishments, but maintains discretion over the exact phrasing and order of the questions. This approach assures that certain data will be gathered from all applicants, but does not assure that comparable data will be gathered from all applicants due to differences in phrasings, sequencing of questions, and length of time spent on particular areas. This is true not only among different interviewers, but also with the same interviewer questioning different applicants.

2. The second most popular approach is the "unguided" or "laissez-faire" interview. In this form, interviewers approach each interview as a unique situation in which they are not bound by areas or questions that must be covered, but are free to pursue those areas that seem to be of most interest. Such an approach allows for a high degree of spontaneity and provides a high level of interviewer motivation. This approach, however, suffers from a lack of standardization since the same questions will not necessarily be asked all applicants for the same position.

3. The least popular approach in employee selection is the structured interview in which the interviewer poses a specific set of questions and deviates as little as possible from the list. Such a procedure is described by interviewers as being

repetitive and monotonous, though it does result in a high degree of standardization, thus increasing validity and reliability of the interview (Schwab & Henneman, 1969; Carlson, et al., 1971; Bass, 1951).

Administration

In addition to the format, the interview can differ according to the number of persons conducting it. It is common for one interviewer alone to conduct the entire interview with a given applicant. However, two other approaches are multiple interviews and group interviews.

1. In multiple interviews, a number of independent interviews with the same applicant are conducted by different interviewers. Topics may be divided among interviewers so that repetition is held to a minimum. Following the entire series of interviews, the interviewers meet together to pool their information and opinions and to arrive at a group consensus. With such an approach interviewer biases or irrelevant data are less likely to affect the final decision. Additionally, gaps in information gathered by one interviewer may have been covered by other interviewers. Thus, a more complete picture of the applicant's background and characteristics is more likely to emerge.

2. A group interview requiring two or more interviewers to jointly question or be present during the interview of a single applicant is relatively uncommon. Interviewers may take turns asking questions or one interviewer may question the applicant,

while remaining group members observe. As with the multiple interview procedure, interviewers meet afterward to make a selection decision. This process insures that all members of the group have the same information on which to base a final decision. With this approach, interviewer biases and irrelevant data are less likely to affect the final decision than with the more common approach of one interview per applicant.

Length of Interview

Despite the fact that the interview is used as a selection device by most organizations, there has been little attention paid to the optimal length of the interview. There has been wide variation reported, ranging from less than 15 minutes to two hours or more. Springbett (1958) discovered that the average interview lasted about 15 minutes. The average interviewer was prepared to make a decision about selection after only four minutes, while certainly, the length of time spent in the interview would vary according to level of position vacancy, amount of ancillary data available (e.g., application blanks, psychometric test data, recommendations), number of qualified applicants, and the purpose of the interview (preliminary screen or actual selection tool).

Instrument Focus

As might be expected from the wide usage of the interview as a selection instrument, the interview has been used to tap numerous applicant characteristics. A review by Wagner (1949) reported 96 different traits and dimensions that researchers have attempted to evaluate by the interview method. Of these, it appeared that the interview was most often used to evaluate overall ability, physical appearance, manner, intelligence and mental ability, judgment and voice quality. More recent studies by researchers have urged that the interview be used instead to measure interpersonal relations, sociability and likeability (Otis, Campbell & Prien, 1962; Holt, 1958; Loevinger, 1959; Ulrich & Trumbo, 1965), and job motivation (Rimland, 1960; Woodworth, 1957) since the highest validities are reported in these areas. The interview is also a good measure of intelligence, but probably less useful than a psychometric test would be to measure mental ability.

Reliability. There is an abundance of information on the reliability of the selection interview. In general, the majority of information has shown high intrarater reliability. When interviewers evaluate the same applicant by replaying the recorded interview or interviewing the applicant again after a period of time has elapsed, they make approximately the same ratings as they did the first time (Shaw, 1952; Pashalian & Cressy, 1953; Anderson, 1954). These data suggest that an

interviewer will approach the interview situation similarly from one time to the next. However, when more than one interviewer evaluates an applicant, reliability falters. In a now classic study, Scott (1915) asked six personnel managers to interview 36 applicants on sales ability. The interviewers not only disagreed on the ranking, but for 28 of the applicants, the interviewers disagreed on whether they should be in the upper or the lower half of the group. Comparable results have been found in numerous other studies (Scott, Bingham & Whipple, 1916; Hollingworth, 1922; Uhrbrock, 1933; Wagner, 1949; Raines & Rohrer, 1955; Plag, 1961; Ulrich & Trumbo, 1965).

One area that has been examined as a source of interviewer error is the temporal positioning of favorable and unfavorable information. Some authors have reported heavier weighting for earlier information, primacy effects, in the evaluation process (Blakeney & MacNaughton, 1971), while other authors have reported heavier weighting of later information or recency effects (Farr, 1973). Peters & Terborg (1975) and Tucker & Rowe (1979) have concluded that a favorable expectancy of the interviewer at the start of the interview followed by negative information will result in more favorable ratings than the negative expectancy followed by positive information. Tucker & Rowe explain this phenomenon by suggesting that when favorable expectancies exist in the interviewer's mind, he or she will give the applicant less credit for personal successes and hold the applicant more personally responsible for past failures.

Contrast effects, resulting from the comparison of job applicants with preceding applicants may also contribute to interviewer error (Carlson, 1969; Hakel, Ohensorge & Dunnette, 1970; Wexley et al., 1972). However, in other studies, contrast effects were shown to be minimal (Hakel et al., 1970; Landy & Bates, 1973).

In recent years, many investigations have focused on how reliabilities can be increased. One such mechanism for heightening interrater reliability is by providing the interviewer with more information about the job to be filled. Langdale & Weitz (1973) reported that interviewers who were provided with job information about the position they were filling, had high interrater reliability (r=.87) while those provided with only a job title had low interrater agreement. A second means of improving reliability is interview structure. Schwab & Henemann (1969) found highest interrater reliability when interviewers used a structured interview format. When a semi-structured format was used, reliability dropped (r=.43), and when an unstructured format was used, the reliability coefficient dropped even further (r=.36). Similar results were shown by Carlson, Schwab & Heneman (1970).

The type of rating scale used has been shown to affect interrater reliability. Maas (1965) conducted a study in which interviewers assessed job applicants for a particular job on two different rating scales. In the first study, interviewers rated applicants on a traditional adjectival rating scale. The reliabilities were low for traits (r=.35), overall rating,

(r=.34) and total ratings (r=.34). When candidates were assessed using a scaled expectation rating scale, reliabilities increased to .58 for traits, .47 for overall rating, and .55 for total ratings. A later replication showed similar results.

Interviewer training has received the most attention as a means of increasing reliability. Numerous studies have shown that training reduces perceptual biases of evaluators (Brown, 1968; Wright, Carter & Fowler, 1967; Borman, 1975). As little as five minutes of training was found by Borman (1975) to be effective in reducing perceptual errors. Shick (1973) likewise reported fewer errors in perception among raters exposed briefly to training.

The best training pedagogy is still under investigation. Levine & Butler (1952) found only group discussion to be effective in reducing "halo" error (a perceptual error in which a wide variety of positive attributes are ascribed to the applicant on the basis of limited positive data). According to Levine & Butler, the lecture method and experience had no effect in reducing the "halo effect." Brown (1968) also examined a variety of training methods and concluded that all methods were effective in reducing the halo error. Finally, Wexley, et al., (1973), and Latham, et al., (1975), found workshops with exercises had the most dramatic effect in reducing rating errors.

Validity

The validity of the interview has been assessed against such criteria as performance ratings, success in training, and job tenure. The results of most of these studies have been disappointing. In his review of the studies examining selection interviewing, Mayfield (1964) stated, "Although the reliabilities of interview may be high in given situations, the validities obtained are usually of low magnitude....This indicates that along with the present emphasis on reliability, there should be more investigation of just what it is that is being measured reliably in selection interviews." In Wagner's (1949) review of 106 studies, the mean validity coefficient reported for traits and characteristics was only .37 and the validity of the overall ratings was .35. It must be kept in mind, however, that the number of studies which assessed validity was quite small and in those studies a number of different criteria were used. Nonetheless, they serve to underscore the generally low validity of the interview as a selection tool. This state of affairs will probably remain the case as long as the interviewers are permitted to draw their own inferences unsystematically from the data they collect.

Most studies that have examined validity of the selection interview have used job performance ratings as the criterion (Raines & Rohrer, 1955; Zaccaria, et al., 1956; Woodworth, et al., 1957; Trites, 1960; Campbell, et al., 1962; Huse, 1962). Other criteria have included the successful

completion of training (Plag, 1961; Trankell, 1959) and job tenure (Hunt, 1957; Williams & Zimmerman, 1957). Regardless of the criteria used, validities for the selection interview remain disappointingly low.

Rundquist (1947) was able to achieve a somewhat higher validity by restricting the scope of the interview to assessing only "social interaction." This study led Ulrich & Trumbo (1965) to suggest, "It may well be that the interviewer is all too frequently asked to do the impossible because of limitations on the time, information, or both, available to him." An approach such as Rundquist's might well allow the interviewer to raise the validity of the interview by focusing on assessment of one area rather than many. The main area which appears to be most accurately assessed by the interview is that of sociability or interpersonal relations. Otis, Campbell & Prien (1962) concluded that the interview yielded valid predictions only on the personal relations dimension. Other evidence in support of their conclusion is found in the studies of Holt (1958) and Loevinger (1959).

A number of other factors also have a bearing on the validity achieved through the use of the interview. One cause for low validity may be errors in information processing, an area which has been exposed to much scrutiny in the last decade. Springbett (1958) and Bolster & Springbett (1961) found that the selection interview was used primarily to access negative information about the candidate. They concluded that though often unintended, the interviewer comes to weight

negative information too heavily (vis-a-vis its actual job relevance) when making an employment decision. In later research Hollman (1972) discovered that negative data are not more potent because they are given excessive inappropriate attention by decision-makers, but because positive information is overlooked or underutilized in the actual decision-making process. The later research would suggest that interviewers should be alerted or taught to attend to and use positive interview data more carefully in order to improve the validity of their interview decisions.

The effect of training on validity has received some attention. In research by Borman (1975) it was found that training programs designed to reduce "halo error" did reduce halo but left validity unchanged. There was some indication, however, that relative strengths and weaknesses of the applicants were recognized somewhat more accurately after training.

Interview structure and the use of biographical information were examined by Heneman, et al., 1975. It was discovered that neither the degree of interview structure nor the use of biographical data influenced validity. The authors speculate that individual judges may have been making decisions based on their own unique stereotypes of the characteristics needed to adequately fill the job. This occurred because of the lack of adequate information about desirable characteristics and behaviors of job incumbents. Indeed, as a study by Wiener & Schneiderman (1974) demonstrated, when information about the job vacancy was

supplied to the interviewer, more relevant and less irrelevant information was used in the selection decision.

Sydiaha's (1961) research suggested there is an "ideal" applicant stereotype against which job applicants are judged. He further felt this "ideal" stereotype was common to all interviewers. However, in a later study, Hakel, Hollmann & Dunnette (1970) found that ideal stereotypes do appear to exist, but they are at least partially unique to the interviewer, rather than being common to all interviewers.

Various authors have expanded on the stereotype hypothesis to examine sex stereotypes which affect interviewer decisions (Cecil, et al., 1973; Shaw, 1972; Heneman, 1977; Cohen & Bunker, 1975). Mayer & Bell (1975) examined sex stereotypes and found that different stereotypes of men and women are responsible for different hiring decisions. Also, the authors revealed that the sex of the interviewer plays a key role as well. Female interviewers had more similar and less complex stereotypes of men and women than did male interviewers.

Table 1 summarizes the reliability and validity of data obtained in the use of the selection interview and documented by the studies cited above.

Issues for Users

Perhaps the most crucial problem facing users of the interview as a selection device is its typically low validity. While studies have indicated that reliability can be raised

TABLE 1

Dimensions Measured by the Selection Interview

Dimension	Reliability	Validity
Traits and Characteristics		
Range	.15-.98	.17-.71
Mean	.56	.23
Intelligence		
Range	.62-.96	.09-.94
Mean	.82	.67
Interpersonal Relations		
Range	.38-.87	.22-.65
Mean	.71	.40
Overall Effectiveness		
Range	-.20-.85	.22-.87
Mean	.45	.41

through techniques such as interviewer training, structured interview format, use of job information and more sophisticated rating scales, the validity of the interview remains low under most conditions.

The only ways which have been shown effective in increasing the interview's validity were restriction of data collected from the interview to interpersonal relations and the use of greater job information which is behaviorally based. Heneman, et al., (1975) stated the dilemma following a sophisticated study on interview validity. "A strong effort was made to create an interviewing process that would result in valid assessments in the structured interview condition." These efforts included (1) thorough job analysis, resulting in descriptions of basic job elements; (2) use of the job elements in the criterion measures; (3) use of interview rating forms requiring interviewers to make explicit predictions of performance on the job elements; and (4) development of structured interview questions directly from descriptions of the job elements. In spite of these efforts, interviewer validity remained low. It is thus necessary to ask what more could reasonably be done in actual interview settings.

The solution to the dilemma lies in a two-pronged effort that would first minimize interviewer error and bias, thus raising reliability. Heightened reliability increases the possibility that validity could be raised. The second effort should be to obtain job competency analyses of positions to be

filled. This information should be supplied to interviewers who would then interview candidates only for those positions where interpersonal relations were an essential skill to successful job performance.

A related question is that of functional utility. Though the validity of the interview is low, it may still be an acceptable component of the selection process, provided it makes a unique contribution to the data needed for a selection decision. However, it appears that the data most easily and validly assessed by the interview may also be assessed, perhaps more effectively, by other methods such as psychometric testing or thought samples. Some researchers (Huse, 1962; Plag, 1961; Ulrich & Trumbo, 1965; Grant & Bray, 1971; Wright, 1969; Schmitt, 1976) have already raised this issue, but relatively little work in this area has been conducted.

An auxiliary question to that of relative utility is the issue of cost-effectiveness. Even if the interview can be shown to make a unique contribution to prediction of job success, tenure or training success, is the gain in predictive ability worth the cost involved?

A third issue which must be addressed is the legal position of the interview as a selection device. The interview has not come under attack to the same extent as other selection devices under Title VII, such as psychometric tests. However, employers who use selection interviews may be called upon to give evidence not only for the validity of the instrument, but also for its

fairness to minority groups and women. As mentioned earlier, it is clear that sex and race do impact interview decisions in quite complex ways. This evidence combined with the typical low validities and reliabilities associated with the selection interview are likely to make defense of the interview quite difficult.

Psychometric Tests

Testing as a tool for personnel selection has been in existence for more than 50 years, receiving its first widespread use during World War I as a selection device for the U.S. military. The development of tests for selection expanded during the following decades, reaching a high degree of sophistication in the post-World War II period. Acceptance has not been as widespread as one might assume. Ward (1960) surveyed 1610 managers on the use of tests in their companies. Of these, 42% reported tests were used for hourly employees while 53% said tests were used for exempt employees. Campbell, et al., (1970) estimated that 60-70% of the companies in the U.S. use ability or aptitude tests, with many fewer firms using personality tests.

There appears to be wider use of tests for selection decisions involving hiring persons from outside the firm. Ward (1960) found about 53% of companies surveyed used testing for external selection, while 36% used testing for internal selection. This differential testing emphasis was supported by

Campbell, et al., (1970) who reported that approximately 90% of companies sampled used tests with external hires; only 40% tested for internal decisions. The rationale given for this difference was that testing is really helpful "only when little is known about the individual."

Construction of Tests

Dunnette (1966) has noted three basic methods for construction of tests: (1) armchair theoretic, (2) factor analytic, and (3) empirical.

1. The "armchair theoretic approach" involves devising a set of materials or questions that will be used to elicit responses from persons, deciding what the responses mean, and then either confirming or disconfirming the actual behavior of those persons tested. This approach is used quite commonly, but has little to recommend it since validities with job performance tend to be low.

2. The factor-analytic approach involves describing and rating actual behaviors shown by persons. These descriptions are then correlated and factor analyzed in order to yield basic dimensions of behavior. Tests are then constructed to measure these various factors.

3. Finally, the empirical approach involves observation of differences in a particular behavior, rating or categorizing individuals according to the amount of the behavior they show,

constructing stimuli which appear to be related to the behavior under study, and then testing the stimuli to see which measures actually differentiate groups of persons having differing amounts of the behavior studied.

Of these three methods, the most desirable is the empirical method because it is so strongly behaviorally based and tends to demonstrate the greatest validity.

Administration and Content

Tests differ substantially in their administration and content areas. For example, tests may be closely timed (speed test) or untimed (power test); they may involve "hands-on" manipulation (performance test), oral questions, or paper and pencil measures; they may have correct answers (objective tests) or no correct answers (subjective tests).

The most important dimension on which tests differ is the content they are designed to measure. Accordingly, the hundreds of tests in use in selection decisions may be grouped under two major headings: ability and skill tests and personality tests.

Ability and Skill Tests

Ability and skill tests are theoretically different, the ability test purporting to measure potential as opposed to level of acquired skill. In practice, the main difference between the two tests is the purpose of the testing: the same test is frequently used to assess potential or actual skill.

Ability and skill tests differ according to the content they measure as well as the specificity of the dimensions tapped. General ability tests usually measure general intellectual ability, while specific ability and skill tests measure particular facets of intellectual, perceptual, psychomotor or other abilities.

General Intellectual Ability Tests

Variety in Format

There has been an evolution of general intellectual ability tests, beginning with the early, "spiral omnibus" tests. These tests assumed a general aptitude which would be measured to some extent by all questions. The result of this type of measure was the single score derived from a number of different types of items. The spiral omnibus variety gave way to a different type of test which assumed a certain number of factors or dimensions comprising intelligence, rather than a single factor. In some cases, an overall score was given as well. Doppelt (1954) reported a growing trend for the factorial concept and a decline among single score tests.

Usually, for personnel selection, intellectual ability tests are administered to groups of persons rather than to one individual at a time. This allows the cost of testing to be kept substantially lower per applicant than would otherwise be the case. In addition, the time required for such tests is shorter than in other settings, again allowing a lowered cost

per person. This need for group administered, short tests of intellectual ability has produced a wide variety of occupational tests designed to measure general intellectual ability. Noticeably absent from the test used in industry is the Wechsler Adult Intelligence Scale, widely used in other settings, but which requires individual administration combined with a rather long completion time of approximately one hour. In contrast, measures commonly used in industry average only 20-30 minutes for a complete administration and may be given to large groups if needed. They tend to be multiple choice speed tests which increase in difficulty. General intellectual ability tests are usually used as a preliminary screening device to be followed by performance or skill tests.

Instrument Focus

Some of the more popular tests designed to measure general intellectual ability are listed in Table 2. As seen in that table, there is a pronounced emphasis on cognitive abilities with some attention to problem-solving ability. In particular, such dimensions measured as fluency with numbers and words, discovering relationships among words and general reasoning are tapped by most of the tests commonly used. However, dimensions such as flexibility, creativity, and problem diagnosis, also part of general intellectual level, are not measured by the tests. Guion (1965) stated, "...the general intelligence tests have been conducted less than general because they do not

TABLE 2

General Measures of Intellectual Ability Commonly Used in Industry

Name	Construction	Reliability	Validity	Dimensions Measured
Otis Self-Administering Tests of Mental Ability S-A	Spiral omnibus-single score obtained Paper & pencil speed test 30 minutes 4 alternative forms			Number fluency Knowledge & meaning of words Classifying verbal concepts Perceiving relationships among verbal concepts General reasoning Spatial relation Ability to reason logically Perceiving events or concepts in logical order
Wonderlic Personnel Test	Spiral omnibus-single score obtained Paper & pencil speed test 50 items 12 minutes 9 alternative forms	Test-retest Range .82-.94 Split-half .88-.94		Number fluency Knowledge & meaning of words Classifying verbal concepts Perceiving relationships among verbal concepts General reasoning Spatial relation Ability to reason logically Perceiving events or concepts in logical order
Adaptability Test	Factor-analytic Multiple scores obtained Paper & pencil power test Spiral omnibus 2 alternative forms	.80	.73-.79	Number fluency Knowledge & meaning of words Classifying verbal concepts Perceiving relationships among verbal concepts General reasoning Perceiving patterns in geometric or verbal stimuli Arranging events or concepts in logical order

(Table 2, continued)

Name	Construction	Reliability	Validity	Dimensions Measured
Thurstone Test of Mental Alertness	Factor-analytic Paper & pencil speed test 126 items 20 minutes	Split-half .95		Number fluency Knowledge & meaning of words Classifying verbal concepts General reasoning Perceiving patterns in geometric or verbal stimuli
Wesman Personnel Classification Test	Paper & pencil speed 60 items 28 minutes	Alternate form .73-.92 Split-half .82-.94		Number fluency Discovering relationships among verbal concepts
Ghiselli Analysis of Relationships	Paper & pencil power test 40 items	Odd-even .82	Concurrent .22-.76	Number fluency Knowledge & meaning of words Classifying verbal concepts Discovering relationships among verbal concepts General reasoning Ability to reason logically

measure the important intellectual powers involved in creative thought, planning and judgment. They do not even tap some of the lower level intellectual skills measured in tests of clerical or mechanical aptitude."

Psychometric Properties

Nonetheless, some of the general intelligence tests have proven to be useful additions to testing batteries for some jobs. Ghiselli (1973) examined all the validity studies (1921-1971) using occupational samples. From these data, he completed the average validity coefficients for the criteria of training and job proficiency. Unfortunately, in computing these coefficients, concurrent and predictive validity studies using all general intellectual ability measures were combined, as were both well constructed and poorly designed studies. The validity figures attained, therefore, are likely to be underestimates of true validity. These validity coefficients which may be seen in Table 3 reveal some interesting trends. (1) General intelligence tests appear to be much better predictors of success in training than of job performance. This is to be expected given the scholastic nature of most intelligence tests. (2) The utility of general intelligence tests is most apparent in clerical and managerial occupations. Overall, validities of general intelligence tests are low (between .15 and .30) for all occupations.

TABLE 3

Validity of General Intelligence Tests*

Occupation	Criterion	
	Training	Job Proficiency
Managerial occupations	.29	.29
Clerical occupations	.46	.30
Sales occupations	—	.19
Protective occupations (e.g., firemen, police)	.65	.23
Service occupations	.42	.26
Vehicle operators	.21	.15
Trades and crafts	.41	.25
Industrial workers	.38	.20

*Adapted from Ghiselli (1973)

Specific Aptitude Tests

Specific aptitude tests attempt to measure an applicant's potential on a specified set of traits and abilities. Of course, in reality, true tests of potential do not exist, but are always influenced to some degree by such factors as previous experience and learning or motivation. Specific aptitude tests may be combined to form a multiaptitude battery with each subtest designed to measure different dimensions of potential for a given job.

Instrument Focus

Specific aptitude and skill tests exist for a variety of dimensions, but in personnel selection the most prominent areas are:

- Specific intellectual abilities including the ability to deal with verbal and numerical materials and geometric forms. Perceptual abilities and memory are often important aspects as well. Specific intellectual ability tests have been most often used to assess clerical aptitude.
- Mechanical aptitude measuring spatial orientations or visualization. These tests may be paper and pencil measures or require actual object manipulation. These skills are most often measured for trades and craftspeople, such as electrical or structural workers, or some machine operators.

- Creativity and judgment which explores some of the higher intellectual processes omitted in general intelligence tests. The traits tapped by creativity and judgment tests include the ability to make inferences, recognize assumptions, deduce logical conclusions, and evaluate arguments. They have been most used in managerial and scientific professions.
- Sensory and perceptual capabilities including such aspects as vision acuity, depth perception, and auditory acuity. Such traits are especially important to clerical workers, inspection workers, vehicle operators, machine operators, some laborers, and mechanic and skilled tradespeople.
- Psychomotor ability including dexterity, eye-hand coordination and object manipulation. These abilities are most needed in work requiring speed and object manipulation, such as assembly of small components or electrical wiring.

Table 4 lists some of the more common tests in each of these categories as well as the dimensions of facets of behavior measured by each.

Psychometric Properties

Specific Aptitude Tests. The average validities of specific ability tests for eight job families were presented by Ghiselli

TABLE 4

Specific Aptitude Tests Commonly Used in Industry

Name	Construction	Reliability	Validity	Dimensions Measured
Minnesota Clerical Test	Paper & pencil speed test	Test-retest .83-.91	Concurrent .28-.64	Perceptual speed and accuracy
General Clerical Test	Paper & pencil speed test	Test-retest .59-.88		Clerical aptitude Numerical ability Verbal ability
Short Employment Test	Paper & pencil speed test 3 five minute tests 1 minute warm-up 4 alternative forms	Alternate form .77-.91 Test-retest .77-.84	Predictive .23-.45 Concurrent .33-.53	Verbal comprehension Numerical ability Memory Perceptual speed
Minnesota Paper Form Board Test	Factor-analytic Paper & pencil speed test 64 items 20 minutes			Spatial orientation Visualization
Survey of Space Relations Ability	Paper & pencil speed test 32 items 15 minutes	Test-retest .93	Predictive .34-.62	Spatial orientation
Christensen-Guilford Fluency Tests	Paper & pencil			Verbal fluency Creativity
Owens' Creativity Test for Machine Design	Empirical Paper & pencil speed test 15 problems 150 minutes	Odd-even .83 Split-half .95		Creativity

(Table 4, continued)

Name	Construction	Reliability	Validity	Dimensions Measured
Watson-Gleser Critical Thinking Appraisal	Paper & pencil power test 8 tests - 230 items No time limit 2 alternative forms	Split-half .50-.84		Inference Recognition of assumptions Deduction Interpretation Evaluation of arguments
MacQuarrie Test for Mechanical Ability	Paper & pencil 7 subtests	.70-.89		Spatial relations Controlled manual movement Visual inspection
O'Connor Finger & Tweezer Dexterity Tests	Object manipulation speed test	Test-retest .89-.93		Manual dexterity
Purdue Pegboard	Object manipulation	Split-half .82-.91	.07-.76	Manual dexterity
Minnesota Rate of Manipulation Test	Object manipulation			Manual dexterity Finger dexterity Wrist-finger speed Positioning

(1973). These may be seen in Table 5. As with general intelligence tests, it appears as if training success is more accurately predicted by aptitude tests than job proficiency. Unfortunately, in Ghiselli's work, he did not report validities of measures of creativity with training or job performance. However, what data there are tend to be concurrent studies which show reasonably good discrimination among creative and non-creative engineers and programmers (Owens, et al., 1957; Langmuir & Kendall, 1961; McNamara & Hughes, 1961).

Multiaptitude Test Batteries. Test batteries which assess aptitude in a number of areas have increased in popularity in recent years. The advantages of such batteries are greater efficiency in the use of testing time, and a greater amount of data obtained about the applicant. Such batteries may be used to assess the wide variety of skills which may be needed to successfully perform the various facets and tasks of a particular job. The use of multiaptitude test batteries has been most prevalent in the military though recently private industry has begun to accept and use some batteries. Table 6 summarizes data from four major multiaptitude tests in current use.

Personality Tests

Variety in Format

Personality tests are designed to measure the emotional, interpersonal, attitudinal and motivational facets of an applicant. Personality measures originated in guidance centers and

TABLE 5

Validity of Specific Aptitude Tests*

Occupation	Ability	Criterion	
		Training	Job Proficiency
Managerial occupations	Intellectual	.30	.27
	Spatial/Mechanical	.28	.22
	Perceptual	.23	.25
	Psychomotor	.02	.14
Clerical occupations	Intellectual	.47	.28
	Spatial/mechanical	.34	.17
	Perceptual	.40	.29
	Psychomotor	.14	.16
Sales occupations	Intellectual	—	.19
	Spatial/mechanical	—	.18
	Perceptual	—	.04
	Psychomotor	—	.12
Protective occupations	Intellectual	.42	.22
	Spatial/mechanical	.35	.18
	Perceptual	.30	.21
	Psychomotor	—	.14
Service occupations	Intellectual	.42	.27
	Spatial/mechanical	.31	.13
	Perceptual	.25	.10
	Psychomotor	.21	.15
Vehicle operators	Intellectual	.18	.16
	Spatial/mechanical	.31	.20
	Perceptual	.09	.17
	Psychomotor	.31	.25
Trades & crafts	Intellectual	.41	.25
	Spatial/mechanical	.41	.23
	Perceptual	.35	.24
	Psychomotor	.20	.19
Industrial workers	Intellectual	.38	.20
	Spatial/mechanical	.40	.20
	Perceptual	.20	.20
	Psychomotor	.28	.22

* Adapted from Ghiselli (1973)

TABLE 6

Multiaptitude Test Batteries

Name	Construction	Reliability	Validity	Dimensions Measured
Differential Aptitude Tests	Paper & pencil power test 4 hours 2 Alternative forms	Alternate form .73-.94 Split-level .96-.99	Predictive -.23 +.23	Verbal reasoning Numerical ability Abstract reasoning Space relations Mechanical reasoning Clerical speed & accuracy
Flanagan Aptitude Classification Tests	Paper & pencil speed & power tests 19 tests 10½ hours	Split-half .65-.86 Alternate form .55-.85	Predictive .04-.65	Verbal fluency Numerical ability Judgment & comprehension Inspection Coding Memory Precision Scale reading Pattern reproduction Mechanical reasoning Mechanical component Table reading Coordination Assembly
General Aptitude Test Battery	Paper & pencil 12 tests			Verbal fluency Numerical ability Finger Dexterity Manual Dexterity General intelligence Spatial relations Form perception Clerical perception Coordination
Employee Aptitude Survey	Paper & pencil 10 tests 55 minutes 2 Alternative forms	Alternate form .60-.70 Test-retest .76-.84		Verbal comprehension Numerical ability Verbal reasoning Numerical reasoning Word fluency Spatial relations Visual pursuit Visual speed & accuracy Symbolic reasoning

mental hospitals, then moved into industrial personnel selection on the assumption that personality is an important determinant of job performance, job tenure and absenteeism, and general quality of life. The number of personality tests is now in the thousands, making test selection frequently difficult.

There are two basic forms of personality measures: the self-report inventory and projective techniques. The self-report inventory requires the applicant to indicate how descriptive statements or adjectives relate to himself or herself. The primary difficulty inherent in this approach is the opportunity for "faking" or giving only socially desirable answers. Particularly in personnel selection, there may be a strong desire on the part of the applicant to "look as good as possible" in order to be selected for the position.

Several procedures have been tried to decrease or eliminate the effect of the social desirability bias in the use of self-report measures. One solution is the use of "forced-choice" measures in which the applicant chooses between two or three possible descriptors, all keyed to have the same desirability. There is some evidence to suggest this minimizes the effect of the "faking", but still does not eliminate it as a source of bias (Wiggins, 1966). A second way to counter the problem of "faking" is the construction of special keys which assess the degree to which responses appear to be heavily influenced by wanting to appear in a more favorable light. These keys also appear to reduce the effect of the bias, but do not eliminate it (Cofer, et al., 1949).

Other difficulties which beset the use of self-report inventories is that of response sets of styles. Inventories lend themselves to such biasing response patterns as acquiescence (the tendency of the applicant to agree with statements) or deviance (the tendency to give unusual or uncommon responses).

A second form of personality measure is that of projective tests. The primary characteristic of projective technique is the use of ambiguous or unstructured stimuli. It is assumed that by using ambiguous stimuli, the applicant is free to project his or her own desires, emotions, needs onto the stimuli, and structure the situation according to fundamental aspects of personal psychological functioning.

Projective techniques are disguised in that the applicant seldom is aware of the interpretation that will be made of particular responses. Thus, faking and response sets are not a problem with this type of technique. Unfortunately, most projective techniques, with the exception of those utilized to collect thought samples, lack standardization of administration and scoring. This makes results obtained suspect, due to lower inter- and intrarater reliabilities. Additionally, many of the projective measures use subjective rather than objective scoring techniques. Normative data, as well, are frequently not available, especially on industrial samples.

Interest Measures

Interest measures were devised originally for clinical use in guidance counseling rather than employee selection. These measures are typically self-report instruments requiring a rating of a particular activity, or a preference among several activities.

The two most commonly used interest measures are the Strong-Campbell Interest Inventory (an earlier version being the Strong Vocational Interest Blank), and the Kuder Preference Record. In the SCII, applicant responses are compared with responses of people in various jobs. These interest similarities and dissimilarities are charted for a number of occupations allowing the applicant to gauge how closely his or her interests match interests of people in those occupations. The Kuder Preference Record requires the applicant to indicate preferred activities in a forced-choice format. The results are given in strength of interest traits, such as mechanical, scientific, artistic, etc., rather than in terms of occupations. In both interest measures there are verification scores to detect faking or carelessness. Other interest measures exist, but are seldom used in industrial settings.

Instrument Focus

Personality tests also vary according to the content and purpose of the test. The two dominant forms of content of personality tests are (1) Motives and Traits, and (2) Interests.

Measures of Motives and Traits

A sampling of the many motivational and trait measures used for selection purposes are shown in Table 6. No one test seems to be used most frequently, though projective tests are used less frequently than self-report measures, probably due to the difficult scoring procedures typical of these measures and the low interrater reliabilities. Most of the measures of motives and traits tap rather ambiguous personality dimensions such as dominance, extroversion, stability or masculinity, which may or may not be reflected in behavior on the job. Cronbach (1960) criticized trait measures by challenging the assumptions on which the tests are based. The assumptions according to him are:

- Personalities possess considerable consistency; a person shows the same habitual reactions over a wide range of similar situations.
- For any habit we can find, among people, there is a variation of degrees or amounts of this behavior.
- Personalities have some stability, since the person earning a certain score this year usually has a somewhat similar score next year.

Each of these assumptions is highly questionable. It seems unrealistic to assume persons will react similarly in diverse situations. Rather, traits should be measured in "terms of behavioral tendencies with a defined class of stimulus situations" (Guion, 1965) that most approach the work situation. Secondly, the assumption that variations of traits are great enough among average people to allow such traits to be used as predictors is questionable. Such measures are likely to pick up the highly unusual cases, but in fact, the majority of persons likely to be tested for positions in industry are unlikely to be situated at the extremes on a measure of a particular trait. Finally, personalities show some change over time which is reflected in low test-retest reliabilities on personality trait measures (McClelland, 1980).

Psychometric Properties. Reliabilities and validities for particular motive and trait measures may be seen in Table 7. Further, the validities of these tests for various occupational groups may be seen in Table 8. The results of this table show good validity of these measures for predicting training success of managers and moderate validity for job performance of sales personnel. For other groups, however, validities are quite low.

Reliabilities of measures of motives and traits have long been a problem. Not only have these tests shown lower reliability over time, but with projective techniques in particular, interrater reliability is frequently low (Anastasi, 1968).

TABLE 7

Personality Tests Commonly Used in Industry

Name	Construction	Reliability	Validity	Dimensions Measured
Minnesota Multiphasic Personality Inventory	Empirical Self Report Paper & pencil		Con-current -.33-+.37	Psychological Dysfunction Masculinity-Femininity Social Introversion
California Psychological Inventory	Empirical Self Report Paper & pencil		Con-current .44	Psychological traits & needs Intellectual & interest patterns
Guilford-Zimmerman Temperament Survey	Factor-analytic Self-report Paper & pencil	.52-.92	Con-current -.27-+.28	Psychological traits
Thurstone Temperament Schedule	Factor-analytic Self-report Paper & pencil			Psychological traits
Cattell 16 PF Questionnaire	Self-report Paper & pencil			Psychological traits
Thematic Apperception Test	Projective 20 pictures			Psychological needs & traits
Rotter Incomplete Sentences Blank	Projective 40 items Paper & pencil		-.73-+.70	Psychological adjustment
Miner Motivation to Manage Scale	Projective			
Strong-Campbell Interest Inventory	Empirical Paper & pencil Self-report 280 items	.75-.84	-.24-+.32	Congruence of interest Patterns with job incumbant
Kuder Preference Record			Predictive -.42-+.44 Con-current -.41-+.70	Interest in different activities

TABLE 8

Validity of Personality Tests*

Occupation	Criterion	
	Training	Job Proficiency
Managerial occupations	.53	.22
Clerical occupations	.17	.22
Sales occupations	—	.32
Protective occupations	-.11	.21
Service occupations	—	.21
Vehicle operators	—	.26
Trades & crafts	.16	.24
Industrial workers	—	.26

*Adapted from Ghiselli (1973)

Psychometric Properties. As with motive and trait measures, predictive validities of interest measures for performance of various occupational groups have been shown to be generally low. However, the validity of interest measures to predict success in training for managers is quite high. One way which has been shown to be a particularly effective means of boosting the validity of interest measures is through the use of empirically derived scoring keys (Boyd, 1961; Knauft, 1951; Tiffin & Phelan, 1953).

In contrast to motive and trait measures, interest tests show remarkably good reliability over time. In a study of the SVIB, Strong (1951) conducted follow-up testing with time intervals of 5 to 22 years. He found reliability of interests were .84 after five years, .82 after 10 years, and .75 after a 22-year interval.

Issues For Users

The use of tests in industrial selection appears to have come under a great deal of attack in recent years. Ebel (1977) states that while the attackers are varied, the target of attack is always the same: the alleged lack of validity of the tests. As we have seen earlier, validities of personality tests and general intelligence tests on job performance tend to be especially low for all occupational subgroups. Reasons for this lack of validity have been postulated by a number of researchers and include poor test construction, poor test selection, and differential validity.

Test Construction

Dunnette (1966) has urged a move away from "armchair theorizing" as the basis of test construction to more sophisticated, behaviorally-based approaches. Yet, a number of tests are constructed on the basis of "pet theories" without adequate attention being paid to the actual behaviors indicative of success on the job. Dunnette states, "Most existing behavioral theories have at best doubtful validity, and it is unlikely that any test developer is so omniscient that he or she can accurately intuit what a person's responses to a set of stimuli may mean in terms of later observed behavior."

In addition, most tests have been constructed to be adapted to the widest variety of industrial situations. This feature lowers the ability of a test to predict the unique behaviors necessary for success in a specific job. Instead, tests should be made much more "situationally specific", i.e., constructed on the basis of behaviors which differentiate successful from unsuccessful performance for a particular job.

Test Selection

Some rationale is always needed for the inclusion of particular tests into a selection battery. Unfortunately, according to Bray & Moses (1972), such rationale is frequently based on test availability or on intuition. A study by Parry (1968) illustrates this point. She asked 10 industrial psychologists to estimate the validities of a number of tests widely used in

personnel selection. Her results showed only one person was able to achieve an accurate estimate of validity, with the other psychologists showing a marked tendency to overestimate the validities. This overestimate would encourage the use of widely known tests without assessing the utility of the instrument for the particular job.

As shown earlier, tests predict better for some groups of individuals than others. Thus, inclusion of a personality test in a managerial battery to predict success in a management training program may be a useful addition. However, use of the same test to select police recruits who would successfully complete training at the Police Academy would likely be inappropriate. Therefore, tests must be examined for their utility in predicting a certain outcome for a given occupational group.

Differential Validity

An additional factor in the attack on test validities is the possiblity that tests may be more valid for some groups than others, i.e., differentially valid. A number of studies have claimed such differential validity exists (Bass & Turner, 1973; Kirkpatrick, et al., 1968; Tenopyr, 1967; Bartlett & O'Leary, 1969; Boehm, 1972) while other studies have claimed differential validity is merely an artifact of poor criterion design (Campbell, 1943; Crooks, 1972) or other methodological problems (Schmidt, et al., 1974). A later section of this chapter deals with this issue in greater depth.

In light of the questionable validity of many tests, as well as the recent attention testing has received in court cases, many companies are re-evaluating the utility of maintaining a testing program as part of personnel selection. Where tests are being retained, they appear to be much more behaviorally focused with less attention being given to person ability or general intelligence measures.

Other Selection Devices

This section is devoted to a discussion of the use of application blanks, resumes, recommendations, work samples and simulations, thought samples and detectors of deception. These devices are dealt with as a group due to their less frequent use and the fact that relatively few studies exist that attempt to describe the reliability and validity of these data sources.

Application Blanks

Many organizations judge a job candidate's potential on the basis of background information about that candidate. They reportedly consider biographical data, educational background, work experience and/or performance on a previous job. One of the "easiest" ways to access some of this information is through the use of an application blank. It is easy because the blank can be designed to include questions on exactly those background areas in which an employer is interested. The form, which can

be reproduced cheaply, provides a standard set of questions that can be given to all candidates. Using an application blank, the prospective employer can gather comparable data from a number of candidates without investing a significant amount of time in gathering that information such as would be required in preliminary interviews; a receptionist or secretary can simply ask all applicants to fill one out. Later, candidates can be compared and those most qualified on the basis of the self-report application forms may then be interviewed or hired.

Variety in Format

Most application forms look somewhat alike, though they may vary significantly in the amount of thought that was applied to designing them and the amount of information that they provide. Most include spaces for indicating name, address, age, weight, schools attended and dates of attendance, period of military service, previous jobs held, and names of personal references. More detailed applications may probe these areas for additional job-related information such as college grades, area of academic concentration, favorite subjects, type of military discharge, responsibilities in previous jobs, the number of people supervised and reasons for leaving a job. Still other application blanks ask specifically if the candidate has held a job similar to the one presently applied for, or ask for the candidate's career plans. The more general forms of the application blank are less expensive, and are usually bought in large quantities

from a business form supplier who imprints the organization's name at the top. The custom-made application blank which asks numerous detailed questions gives much more specific job related information at the greater cost incurred by analyzing the specific job to generate questions related to important areas of job functioning. The general form can be used for a variety of jobs; the specific form is only appropriate for jobs that require a particular kind of background and preparation.

While the majority of organizations have all applicants complete application blanks, many of those organizations would be surprised to hear the application blank called a selection device. Most firms use a general form for entry into all areas of the organization. While such forms help the organization to gather data, these data are used in an idiosyncratic manner, unspecified weight being given to that information which may be combined with interview test or reference data. A small minority of organizations use "weighted" application blanks, which are empirically keyed to and validated against job performance. Weighted application blanks are also used as selection devices, either by themselves or in combination with other data.

Instrument Focus

What information do application blanks provide to an employer? At the most basic level, the devices can give basic indications of literacy, spelling, grammar and usage, as well as neatness and the ability to follow simple directions, and they

provide verifiable age, education, military service and employment histories. If accurately recorded, these data provide an uncomplicated profile of applicant skills and job related experience. Questions about one's reasons for leaving a job, or one's career goals, on the other hand, may provide a crude reading on the motivation of the applicant. This information may be helpful in deciding if an applicant would "fit in" on the new job. However, as the self-report information on the application becomes less objective, and therefore less verifiable, it also can be less detectably faked. All self-report knowledge, skills and abilities should probably be verified, using the application blank only as a preliminary screen to identify potential employees rather than for making hiring decisions about jobs that require those knowledges, skills and abilities.

Psychometric Properties

Reliability. The reliability of an application blank, generally speaking, is not at issue. One's education, job history, and the opinions of one's references would not be expected to change very much over short intervals, to vary with changes in the format of an application blank, or to contradict other pieces of information on the same form. There is, nonetheless, the potential problem of candidates faking information in order to present the image they think will get them jobs. Early researchers found self-report data on work history, and

related information to be highly reliable, correlating approximately .94 with verified data (Keating, et al., 1950; Mosel & Cozen, 1952). More recent research suggests that applicant information may disagree with verified information as much as 57% of the time (Goldstein, 1971). However, among incumbents who need not sell themselves for the job (Cascio, 1975), and among applicants led to believe that a faking scale is included in the application blank (Schrader & Osburn, 1977), verified responses indicate that application blank data are highly reliable.

Validity. The validity of biographical items on application blanks, as compared to personality measures and other predictors, has been good for the prediction of job performance (Asher, 1972); biographical items have outperformed intelligence, aptitude, psychomotor, perception and personality scores as predictors of job proficiency. It is important to note, however, that it is a rigorously defined subset of the universe of all types of biographical data which has fared so well in such comparisons. Asher reported and compared only uses of "hard" biographical items, items which have been cross-validated, and which have been combined as a predictor based on a set of biographical questions rather than used as single items. Of 11 studies and 31 validity coefficients meeting this set of criteria, all but 3% achieved validities greater than .30, and 35% of the reported validity coefficients were in

excess of .60. Many employers, however, use data from application blanks which may include "soft" unverifiable items which can be faked easily. Also, many users simply apply a scoring key to an application blank and use the key for selection without ever cross-validating it on another applicant sample to identify the dependence of predictiveness on chance sample variability. Still other users attend to individual items on the application blank, ultimately using one or a few of them as predictors instead of using the combined biographical profile which will, by definition, be able to explain more of the variance in job performance. Asher drew his conclusion about the utility of biographical items from a sample of methodologically sound uses. Users of biographical data (Thayer, 1977; Roach, 1971) reporting validities between .30 and .40 have also argued forcefully for use of application blanks which are continually cross validated and updated to match the job-relatedness of background items for new applicant populations.

Rigor is required to maintain validities even at a moderate level (r=.35). Economic and organizational climate, as well as changing personnel practices, have a continual influence on the predictive validity of information gathered on an application blank. If it is acknowledged that none of these factors is static, then the good predictive validity of biographical data (e.g., personal, education and job history information) from application blanks is something that must be constantly monitored, and changes should be made when necessary to preserve validity.

Utility and validity of the application blank for gathering predictive biographical data has been demonstrated for a variety of jobs types and job environments. Keyed application blanks have been used in the insurance industry since 1922, yielding cross-validated validity coefficients around .45 (Roach, 1971). Typical of the validities obtained for different jobs are .58 referenced to performance in training and on the job for Navy divers (Helmreich, et al., 1973), .36 for men in the Israeli Army (Nevo, 1976), and .67 referenced to quality of performance among dance, theater, music, and visual arts students (James, et al., 1974).

One study of the application blank examined both its reliability and its validity for prediction of job performance. In a crosscultural study of American and Western European salesmen, a biographical inventory demonstrated good median internal consistency, or reliabilities, of .75 and a median validity coefficient of .42 (Hinrichs, et al., 1976). The 48-point range of validities, however, reinforced the author's point that sample variations and variations in personnel practices necessitate validation of the keyed applications from each company and work group despite perceived similarity of job titles and job descriptions between groups. Similarly, the variation in reliability estimates between groups (a 29 point spread) raises questions about whether experienced job applicants are less reliable in their responses, and whether that lower reliability suggests an effort at self-presentation which decreases reliability and thence, validity.

Issues For Users

At this point it should be clear that beyond the basic concerns of reliability and validity, there are several other serious issues to be considered when using an application blank as a selection tool. The issue of fakeability is a serious one; faked responses may reduce both reliability and validity of the method. Cross-validation is also important: a validity coefficient for a sample on which the device is keyed will often be reduced to insignificance when the key is applied to a sample which doesn't have the unique pattern of variance the original sample had. A third issue is one that was passed over earlier; biodata validity may be different for different applicant subgroups divided on the basis of age, sex, race, education, and other background variables. One large trade organization has developed close to 50 different keys to scoring its biodata blank due to changes in applicant populations over time and to differences among subsamples of the population (Thayer, 1977). A fourth issue concerns the nature of published studies. It has been argued (Schwab & Oliver, 1974) that few studies have been published on the validity of applications' biographical data because they are seldom valid predictors of performance. Studies finding no validity, or lost validity upon cross-validation, tend not to be published, but this is difficult to verify.

Lastly, a most serious concern for practitioners today is the legal defensibility of a selection technique as "job related". It has been suggested that data from the application blank be weighted according to their job-relatedness (Pace and Schoenfeldt, 1977) in order for the application to be a job related selection device. This complicates what was originally an empirically simple method. Nevertheless, employers who are committed to using biographical data of the types included in an application blank may be able to meet legal requirements and develop a valid scoring key in this way (Cascio, 1976).

Resumes

Many organizations screen, or even select, applicants on the basis of resumes they submit. A resume is similar to an application blank in terms of the information it supplies to the potential employer. The resume, however, presents a less objective, less standardized self-portrait by a job candidate. Like the application blank it suffers from the typical weaknesses of self-report measures. Applicants present themselves as they wish to be seen, including only what they want and elaborating on information expected to cast them in a positive light. A resume often may not include specific job related information, such as reasons for leaving the previous job and school grades, that the employer might desire.

A resume is prepared by the applicant and given to the prospective employer. It typically includes a personal section, including information on age, marital status, and number of

dependents, memberships and offices held in social and professional organizations. An education section will list the schools attended, years in attendance, and degrees obtained, and may also include grade-point average, academic honors, concentration of study or areas of special interest. A section on job history may simply list organizations and job titles or may be expanded to give varying amounts of detail on responsibilities, descriptions of projects and salary history, and professional applicants may also list job-related publications or public presentations. Clearly, considerable variety in content and specificity may exist among even a small sample of resumes for the same job.

Resumes do provide the employer with biographical information about the job candidate which includes some indication of personal interests, information on the candidate's education, and related job experience. Any self-report information on knowledge, skills and abilities, however, does not necessarily constitute definitive evidence of them. These self-reports may be more or less accurate depending upon the distortions (intentional or unintentional) inevitable in self-presentation. To the best of our knowledge, no evidence of the reliability or validity of resumes exists. However, all those who have written resumes recognize that they are written to specific audiences, for specific jobs and are often rewritten for new audiences and opportunities. This would tend to reduce slightly both reliability and validity of the resume as a data source.

In addition to applicant biases, there are situational and personal biases introduced by the decision maker which will affect the use of resume data. For example, it has been shown that a resume will be evaluated differently depending upon the quality of the preceding resume which serves as an unwitting standard for comparisons (Hakel, et al., 1970). While these contrast effects on resume evaluation are real, they account for a very small portion of the variance (about two percent) in the ultimate interviewer decision. If a candidate is only mediocre in qualifications on a resume, however, being preceded by a terrible candidate may make him or her look good by contrast, and result in an offer of a follow-up interview. Resume evaluations also tend to be more positive for attractive and qualified male candidates (Diploye, et al., 1975, 1977). Content areas, scholastic history, interest, and experience are all important inputs into resume ratings, which tend to be a function of the importance of the particular content area for job performance and the favorability of the information (Hakel, et al., 1970). A systematic review of the biases involved in resume evaluation may be found in Arvey (1979). In sum, there are both job-related and unrelated influences which affect the validity of employment decisions based ostensibly on resumes.

In conclusion, the resume is widely used because of the potential wealth of information it can provide about a candidate's personal, educational and job histories. It is, however, a biased form of data, but one which is used for a first

cut at whom to invite for a later interview. If used as a prescreen for standard selection procedures, the basis for choosing resumes should be demonstrably job-related or else it may be vulnerable to legal contest. Those who use resumes as the basis of selection, or as a prescreen should also be aware of non job-related biases which may affect and invalidate those judgments.

Recommendations

Employers or personnel directors may consider recommendations from acquaintances, other employers or colleagues, or individuals suggested by the applicant, when making decisions about a job candidate. As with application blanks and resumes, recommendations may be used directly to make hiring decisions, but are more likely to be used as the criterion for inviting a candidate to an interview. While some employers say they value a phone call to a past employer or the verbal opinion of a colleague, only recommendations submitted in writing may be studied in any systematic way. Therefore, the present discussion is restricted to written recommendations requested by the employer.

A written recommendation may take a variety of forms. The prospective employer may simply request a certain number of reference letters of the applicants or a form letter may be sent to persons named by the candidate. This letter may vary in structure ranging from a space for candidate and reference names, some simple directions and perhaps guiding questions

specifying length and content of response, followed by a blank space for writing, to a detailed questionnaire including free response, multiple choice, ranking, and forced-choice questions tapping information pertaining to work habits, personality, employment history, and whether the reference person would hire the candidate. Both the more and less structured forms have the advantage of letting the referring person communicate those things about the candidate that he or she knows uniquely and which can't be obtained from an application blank or interview. At the same time, multiple unstructured references for candidates for a single job may vary in content, quality and specificity. Comparison of unequal data and reconciliation of contradictions is difficult under these circumstances. On the other hand, structured recommendation forms provide comparable data, from multiple sources or for multiple candidates, but may not be flexible enough to access the unique data the reference person may have about the candidate.

The exact reason for using recommendations varies by job, company and particular user. In general, recommendations are intended to obtain information on job-related skills, employee character, work habits and employment history. It is the most conventional way to check on what an applicant says he or she has accomplished. A candidate will tell you <u>what</u> jobs he or she held, and what responsibilities he or she had. An employer will tell you <u>how well</u> the candidate fulfilled those job requirements.

Unfortunately, since most recommendations are relatively unstructured, they are consequently somewhat unreliable. Different supervisors and acquaintances have different levels of writing skill, their vocabularies differ, and their skill at person perception and understanding of the new job are highly variable. As a consequence, the data they contribute are also variable, resulting in interrater unreliability. In addition, employers react favorably to well-written recommendations, regardless of the quality of the candidate described. While writing skill may reflect something about the basic intelligence of the author-observer, an employer will have difficulty operating the erudition of the reference from the important qualities possessed by the job applicant. As a consequence, judgments based on reference quality rather than content may be invalid, reflecting more on the writer than the subject of the letter of reference (Mosel and Gohen, 1959). There is also some evidence that employers stereotype authors of recommendations by their sex, and judge reccmmendations accordingly (Kryger & Shikiar, 1979). Nevertheless, recommendations may include valid, useful information on the applicant's character: the validity of such information is best determined for the population and job of interest (Mosel, 1958).

In sum, when recommendations are used as part of an employment process, three issues need to be considered. (1) The employer should recognize that unstructured formats are unreliable. The user should determine whether to structure content

or prescribe reference sources in order to reduce that unreliability. (2) Those who actually use the recommendation in a decision process should be forewarned or trained to attend to content rather than style of the reference, and to consider the related validity issue. (3) With the advent of the 1974 Family Educational Rights of Privacy Act, applicants who submit letters of reference from teachers and professors may request that those letters be placed in a file open to candidate inspection or a confidential file to which the candidate has no access. While it is unclear exactly what the difference would be between two letters by the same author for the two different files, it appears that employers react more favorably to confidential files, regardless of the enclosed recommendations on candidate competence (Shaffer, et al., 1976).

Work Samples/Simulations

One method of determining a candidate's suitability for a job is to have the candidate try his or her hand at a simulation or sample of job tasks. While we may refer to all such tests as work samples, it is important to remember that there are potentially as many different forms of work samples as there are different jobs to be filled. They can be roughly classified into two groups: motor, involving the manipulation of things, and verbal, usually language-oriented or people

oriented (Asher & Sciarrino, 1974).* These work sample assessments are usually conducted outside the personnel office and away from the actual job situation in a place where there will be no interference from factors unrelated to the test and where standardized observation of performance is possible.

Work sample assessments are generally intended to ascertain the level of specific job related skills that the candidate possesses. Self-report of skills, from interviews, resumes or application blanks, can be verified in work samples. Coordination, planning and other cognitive skills, as well as interpersonal and motor skills which may be very important for good job performance can be evaluated by observation of work samples, although they cannot be assessed from most paper and pencil tests. While an applicant may demonstrate that he or she possesses some knowledge by taking a paper and pencil test, a simulation is more appropriate for demonstrating an ability to apply that knowledge.

The reliabilities for work sample tests generally are not reported. It is not unreasonable, however, to conclude that

---

* The verbal category would include most cognitive tasks, for even if the cognitive process involved neither language nor people, the output would have to be verbal to be evaluated.

they must be fairly high on the average on the basis of the similarity between work samples and skill tests. Validities are fairly high for work sample tests, an impossibility without good reliability. Also, work samples are by definition highly structured situations where specific tasks are observed and prescribed actions are credited as appropriate. However, to our knowledge, no useful reliability data exists to support these inferences. Nevertheless, as already suggested, the validity evidence for the work sample is strong in a review of over 60 validity studies for both motor and verbal work sample tests (Asher & Sciarrino, 1974). Two-thirds or more of all the validity studies reviewed exceeded a validity coefficient of .30, regardless of whether the work samples were verbal or motor, or whether the criterion was job proficiency or training success. More specifically, the literature suggests that verbal work samples outperform motor work samples in predicting training success, and motor work samples likewise outperform verbal work samples for prediction of job proficiency. However, Asher & Sciarrino (1974) found that work samples of both kinds consistently finish behind biographical data in validity, referenced to the criterion of job proficiency.

These findings raise some issues, as yet unresolved, about the nature and use of work samples. First, if, as has traditionally been argued, a work sample tends to be a valid

predictor because of its point-to-point relationship to the job performance criterion, then why do biographical data items tend to be more valid? Perhaps the focus on behavioral matching overlooks the facilitiating motivational set which can be conveyed in biographical data. Second, Weitz & Adler (1973) suggest that simulations should be short or of moderate duration. In long simulations, subjects may start to adopt simulation-specific skills which are not job-related, and in fact interfere with transference of the basic work skills to the real work situation. To the best of our knowledge, no one currently using work samples for selection considers this caveat from the history of simulation training. Third, while different racial groups may perform equally well on the work sample, they may have differential attrition rates on the job (Farr, et al., 1973). If turnover is a criterion of interest, then a user should do careful subgroup analyses of work sample performance and subsequent nontask behavior in the work setting.

The last, and most basic issue is, how does one construct and score a work sample? Clearly some kind of job analysis is necessary to build a simulation with content and face validity (Campion, 1972), but a job analysis that simply lists tasks or outcomes is insufficient without a component that ranks or rates their importance for job functioning. It has been shown that a content-valid simulation, the "in-basket" exercise, may

not be predictively valid unless it is scored on the basis of those skills which are most important for the job (Brass & Oldham, 1976).

Thought Samples

A special technique somewhat similar to work sampling has been developed to assess, in particular, motive dispositions related to various jobs. It involves objectively coding samples of a person's thoughts in imaginative stories written in response to pictures. The codes were originally derived (see McClelland, et al., 1953; Atkinson, 1961; Winter, 1973) by identifying what characteristics of thought regularly appeared when a given motive was aroused. Then, if those thoughts occurred in stories written by subjects not under conditions of motive arousal, it was assumed that they were generally under the influence of the motive which uniquely produced such thoughts. In this way, measures have been developed for the need for Achievement (McClelland, et al., 1953), the need for Affiliation (Atkinson, 1958) and the need for Power (Winter, 1973).

The general logic of this approach assumes that being concerned about certain issues, or thinking a lot in terms of certain goals, means that a person will act in ways that are especially appropriate for success in particular jobs. The logic is not unlike that for work samples: what a person does

(thinks) in a sample situation predicts what he will spend his time doing on the job. Two motivational thought patterns in particular have been associated with successful vocational outcomes. The first involves the need for Achievement, or the tendency to think a lot about doing things well or in a more efficient way. Such a thought pattern has been regularly found to be associated with success as a small businessperson or entrepreneur (McClelland, 1961, 1966; McClelland & Winter, 1971; Mum & McClelland, 1979). The relationship has good theoretical validity because it makes sense that a person who thinks a lot about doing better, e.g., getting more output for less input, should be just the kind of person who will succeed in a small business which requires constant attention to input/output ratios.

The second motivational or thought pattern is associated with managerial success in larger businesses. It is called the leadership motive pattern and involves a relatively high need for Power which is higher than the need for Affiliation, and a high sense of self-control. This motive pattern has been found to be related to success as a sales manager (McClelland, 1975), to rated performance as a Naval Commanding and Executive Officer (Winter, 1979) to promotion to higher levels of management within the American Telephone and Telegraph Company over a 16-year period (McClelland, et al., 1980) and generally to success in top management jobs in

American companies (Boyatzis, 1979). Again the relationship has good face validity, for it means people who make good managers tend to think a lot about influencing others, carefully control their influence attempts, and are not excessively concerned whether they are liked or disliked (need for Affiliation).

The reliabilities for measures obtained from objective coding of thought samples have generally been reported to be low (Entwisle, 1972). However, there are important reasons for believing that the coefficients reported may be serious underestimates of the true stability of the measures. As McClelland has pointed out (1971), the instructions for the test tell the subject to "be creative," which is interpreted by them to mean that they should tell <u>different</u> stories each time. It is well known that in all organisms there is a built-in tendency to vary spontaneous responses, which has been called "associative refractory phase." Winter and Stewart (1977) have demonstrated that if the variability set is broken by telling subjects on the second administration of the picture-story test that they are free to tell the same or different stories, then test-retest coefficients rise to the more respectable level of .60. More importantly, the validity studies mentioned above indicate that the measures must have a higher reliability than what has typically been reported, if we assume that validity cannot be higher than reliability.

Finally Atkinson and Birch (1979) have argued that for measures of spontaneous behavior like these, the traditional psychometric model of reliability does not apply.

These measures have an important but limited utility. They are objective in the sense that the coding schemes for them are precise enough for two different judges to obtain a high degree of agreement (r=.85-.90) in coding the same protocol. For types of positions in entrepreneurship and management for which the most research using them has been carried out, they have good face validity, and they do provide information on the motives needed for these two types of jobs which is not obtainable in any other way. On the other hand, their limitations are: (1) that the testing conditions must be carefully controlled since stories written can easily be influenced by situational factors; (2) that they are more costly to score than tests involving machine scoreable choices; (3) that their coverage of competencies and types of jobs is so far quite limited; and (4) that the method is often viewed with some suspicion because it appears to be getting at some unconscious aspect of the self.

Nevertheless, the principles behind thought sampling--objectively coding thought patterns--can be extended to selection techniques such as the interview to determine the response of a wider range of skills. McClelland (1973) and Boyatzis (1979) have reported success in isolating and

reliably coding evidence for intellectual and interpersonal competencies in top performers through unstructured self-report. These competencies were then made the focus of specific psychometric tests which successfully validated their presence in better performing job incumbents.

Detectors of Deception

One group of employment prescreens seldom studied by psychologists are the detectors of deception. These include the polygraph, voice stress analyzer, and paper and pencil methods of detecting deception. Each of these methods is used occasionally as an employment prescreen by organizations concerned with employee theft or confidentiality of corporate information. The polygraph is designed to measure physiological indicators of stress such as pulse rate, relative blood pressure, rate and depth of respiration and galvanic skin response (GSR). Voice stress analyzers presumably measure inaudible stress-related frequency modulations in the voice. Stress is assumed to increase when the applicant is trying to deceive the tester. The paper and pencil instruments are designed to access information about applicant attitudes toward theft, admissions of theft, and biographical correlates of deception. The actual questions included, equipment involved, and time spent on detecting deception may vary between different organizations but all of these employment screening

processes are administered in the employment office or testing center. Administration would appear to be under constant conditions within an organization.

Strictly speaking, detectors of deception do not try to identify any job related knowledge, abilities or skills. If an applicant should try to deceive the employer about his or her background or experience, this would presumably be detected as physiological or vocal stress. The employer, however, appears to focus the examination on issues related to honesty, theft, and lying behavior rather than on issues related to job-related skills.

Psychometric evidence of reliability and validity is tenuous for these employment prescreen methods. Reliabilities in the .80s and .90s have been reported for the polygraph between graduates of the same training program, but reliability drops significantly for more heterogeneous groups of polygraph operators. Likewise, the percentage of accurate response (validity) reported for polygraph use is presentable but the research designs do not correspond to the realities of employment situations (i.e., accuracy is assessed in circumstances when all subjects have some piece of information to hide and accuracy is equal to the hit rate for an operator in identifying the deception). In most employment situations, the operator doesn't even know what areas of the individual's history include deceptions, and very few applicants in the

total applicant population may have some employment-related problem to hide. Demonstrated validity for voice stress analyzers is even lower, and again, is determined in situations not analagous to employment settings. Sackett & Decker (1979) review the reliability and validity data on all three techniques in some detail, and conclude that the techniques are fairly widely used in criminal investigations and employment selection decisions, but the validity data that exist are inadequate to support the use of these methods for selection.

In sum, most of the comments that can be made about the use of detectors of deception are in the form of cautions. Before implementing polygraph tests as employment prescreens, one should be sure to check state law. Use of the polygraph is currently restricted in 15 states, with 19 states requiring licensing for all polygraph operators. There are also ethical considerations in the use of such devices. Is the polygraph an invasion of privacy, as is argued about personality tests? Can one decide that someone is lying on the basis of a test of unknown validity in the employment setting? These two questions bring us to an issue at the heart of any application of detectors of deception: Are these methods valid in an employment setting where base rates of deception are low? Virtually all research on the reliability and validity of detectors of deception has been conducted in actual or simulated criminal investigations. The ability of these data to

be generalized to low base rate employment situations is, as yet, unproven.

Table 9 summarizes reliability and validity data on application blanks, resumes, recommendations, work samples and simulations, thought samples, and detectors of deception. These data are less voluminous than those for the interview and the psychometric test, and are presented together for ease of comparison.

Differential Validity and Test Bias

When considering the merits of using any one of the preceding selection methods, whether from the compliance perspective of a government agency or the pragmatic perspective of an employer, a primary concern is whether their use will result in adverse impact on protected classes of applicants (i.e., subgroups identifiable by race, creed, color, sex, religion, national origin, age, marital status or handicap). We define adverse impact as disproportionate hiring of individuals on the basis of their potential job performance. For example, when a hospital accepts applications and selects chaplains who are male and Catholic only, this does not constitute adverse impact on members of other groups, because being male and being Catholic are bona fide occupational qualifications for a position which requires a person to give the last rites to Catholic patients. Other individuals are excluded on the

TABLE 9

Reliability And Validity Data By Technique

Method Study	Reliability Estimate	Reliability Method	Validity Coefficient	Validity Method	Criterion
APPLICATION BLANKS					
Goldstein (1971)	43% to 85% depending on item	agreement between applicant data and that data verified by checking with past employers			
Cascio (1975)	$.41 \leq r \leq 1.0$, median r= .94	correlation of applicant data and verified data			
*Keating, E. et al (1950)	r=.90 to .98, mdn= .94	correlation of employee and employer data			
Mosel & Cozan (1952)	r=.87-.98, mdn=.94; mean agreement on job duties 85.5%	correlation of applicant and employer data on work history; job duties compared by % agreement			
Roach (1971)			.29	predictive criterion related validity	tenure in months
Asher (1972) (review)			55% of reviewed validity coefficients are greater than or equal to r=.50	criterion related and cross validated	job proficiency

*work history gathered by interview method

(Table 9, continued)

Method Study	Reliability Estimate	Reliability Method	Validity Coefficient	Validity Method	
APPLICATION BLANKS (cont'd)					
Helmreich et al (1973)			.59	cross validated criterion related validation (predictive)	success in Navy diver training
James et al (1974)			.67	criterion related (concurrent); not cross validated	art vs. non-art student
Nevo (1976)			.36 (men); .18 (women)	criterion related cross validated (concurrent)	military rank at discharge
Hinrichs et al (1976)	.78[k] .75[k] .65 .77 .58 .75 .49	KR20 internal consistency within each sample	.72[k] .72[k] .42 .56 .24[ns] .38[ns] .26	(concurrent) criterion validity; cross validated	pooled overall ratings by three executives

*all correlations are significant ($p < .05$ or better) unless otherwise indicated (ns) k-samples used to generate scoring key

RESUMES

NO DATA AVAILABLE

RECOMMENDATIONS

Mosel & Goheen (1958)	4 of 12 validity coefficients for 12 jobs were significant; $.21 \leq r \leq .29$	(predictive) criterion related	job proficiency from supervised performance ratings

89

(Table 9, continued)

Method Study	Reliability Estimate	Reliability Method	Validity Coefficient	Validity Method	Validity Criterion
WORK SAMPLES/SIMULATIONS					
Asher & Sciarrino (1974) (review of 60 studies)			% of $r_<$ exceeding the given value		
			r motor verbal .30 78% 60% .40 70% 41% .50 43% 21%		job proficiency
			r motor verbal .30 79% 81% .40 47% 65% .50 43% 39%		training success
DETECTORS OF DECEPTION					
Sackett & Decker (1979) (review)					
*Polygraph	r=.80	interrater reliability	80-90% accuracy	compared to known base rate of guilt or to expert opinion	base rate or expert opinion
Voice Stress Analyser			median accuracy reported =.30	compared to known truth	known truth
Paper & Pencil forms			median r in low 40's	correlated with theft admissions	theft admissions

*Note: These figures make assumptions of base rate of dishonesty which are unrealistic and inappropriate for polygraph use in employment

basis of their sex and religion but that exclusion is job-related. If, however, a hospital refused to consider or hire female applicants for the position of Protestant chaplain, their actions would have adverse impact on women since both men and women can be ordained ministers in most Protestant sects. The exclusion on the basis of sex would not be job-related.

Adverse impact is the visible consequence of what psychometricians and industrial psychologists call test bias. In this phrase the word is used in the very generic sense adopted by the EEOC and other federal agencies charged with assuring use of fair employment practices; a test is any paper and pencil or other measure used as the basis of an employment decision. In the preceding examples the implicit question, "what sex are you?" is the test used as a basis of the employment decision. The phrase test bias is used to refer to unfair consequences of using tests (predictors) as the basis of selection decisions in a real life situation. A test which is used to disproportionately eliminate applicants on a non-job related basis shows test bias. The most visible and socially unacceptable examples of test bias disqualify minorities, the handicapped, etc., from jobs on the basis of predictors (tests) that are not in fact job-related. Test bias may also occur but result in no adverse impact. For example, as an employer, one might arbitrarily decide to hire

only receptionists who indicate on their applications that they are left-handed. This would consititute test bias since being left handed is not job related. It would not, however, constitute adverse impact since no protected classes would be disproportionately eliminated from consideration; all right handed people regardless of subgroup membership would be disqualified. This kind of test bias is not illegal under Title VII of the Civil Rights Act of 1964.

For the purposes of federal agencies, employers and the law, test bias resulting in adverse impact is a serious problem and the driving issue behind compliance reviews, careful choice of selection procedures, and court cases. Therefore, given the preceding definitions and the concerns of our readers, we will focus now on the potential for illegal test bias (i.e., adverse impact) from use of selection predictor measures.

Since test bias refers to unfair consequences of test use it is not surprising that several different authors have attempted to define what is an unfair consequence. As a result there are now nearly as many different models of test bias, defining unfair consequences in different ways, as these are authors who write about test bias. For example, there are 11 models of test bias included in Peterson & Novick (1976), each of which defines fair use of a test differently. Basically what these models do is define what kinds of

selection errors must be minimized for a test to be fair to all applicants. Some authors say that a test must be used to maximize choice of individuals who will succeed; this is a fair use of a test. Others say a test should be used to minimize the likelihood of either rejecting someone who would have succeeded, or accepting someone who would have failed. Still others try to maximize the chances of rejecting a person who would have failed. Even though these do not sound like contradictory goals, each definition of what constitutes fair use of a test implies different statistical corrections should subgroup performance differ on the predictor. Measures taken to prevent test bias vary greatly depending on the model accepted by an organization. To use a test fairly an organization may choose appropriate cut-off scores and select proportions of individuals from subgroups so as to maximize its goal in terms of fair consequences while minimizing nonjob-related disadvantages to members of subgroups.

The Uniform Guidelines on selection practices, as well as their predecessors, have focused on adverse impact and in particular on differential validity leading to adverse impact. The term differential validity refers to a situation where the correlation coefficient between a selection procedure (e.g., test score or information used as a predictor) and job performance is significantly different for different applicant subgroups. In cases where a difference in these

validity coefficients exists between subgroups identifiable by race, sex, etc., but the test score or predictor information is used in the same way to select applicants without regard to subgroup differences, test bias and adverse impact results. An organization that does not closely examine its selection procedures for differential validity may end up using its selection procedures in ways which have unfair consequences. However, once a model of test bias has been chosen which targets unfair consequences of test use, statistical corrections can be made and selection cut-off scores set to reduce or eliminate adverse impact from use of the predictor test. Therefore, though differential validity can obviously lead to adverse impact, it need not do so. The potential unfair consequences of differential validity can be statistically eliminated while still allowing for use of the test.

The explicit attention given to differential validity as a cause of adverse impact has blinded many users to the broader issue of unfair test use.(i.e., test bias resulting in adverse impact). It is very important to note that an absence of differential validity does not preclude adverse impact. Consequently, an organization which assumes it is using a test fairly because correlations between tests (or predictors) and performance are equal across subgroups may be in error. If the range or means of predictor scores for

those subgroups differ, setting predictor score cutoffs to select applicants as if the groups were identical will result in test bias; selection without attention to characteristics of the subsample scores (including mean, range or validity coefficient) may also result in adverse impact.

Adverse impact, therefore, results from test bias and may be the consequence of any number of factors, of which differential validity is but one. Thus, as suggested by Fincher (1975) and Schmidt, Berner and Hunter (1973), differential validity may be a pseudo problem that overshadows the problem of test fairness.

Early court cases concerned with discrimination resulting from use of employment tests focused on test bias and differential validity without clearly distinguishing between the two. In the classic Griggs case the courts disallowed the Duke Power Company's test because they were not demonstrably job-related for the blacks who were required to take them to advance, and because they were used in a way that had clearly negative consequences for the black workers. The impression gathered from our sample of employers is that organizations have interpreted the ensuing court cases and public policy as being solely concerned with identifying cases of differential validity even though, as indicated above, differential validity is not necessary for test bias or adverse impact. Nonetheless, policy-makers in business and government have

been rightly concerned with the need to conduct subgroup analyses and be wary of the adverse impact that may result if a generalized decision rule for hiring is used to select applicants from sub-groups whose validity coefficients differ. Corporate experience and academic research have demonstrated that differential validity sometimes occurs.* More important issues are: when and why differential validity occurs; how the controversy over the existence of differential validity relates to the central issue of test bias; and what steps are possible to prevent adverse impact of test use.

We refer to recent writings and opinion on differential validity as a controversy because the current trend among researchers is to aggregate individual studies of differential validity and draw conclusions about whether differential validity in fact exists at all! Researchers and practitioners who have calculated subgroup validities and found significantly different validity coefficients may be very skeptical of any research which questions the existence of differential validity. Nonetheless, a few researchers have examined large numbers of studies in an attempt to determine if findings of

---

* For a compete review and bibliography of all such published studies of differential validity, see reviews by: Boehm, 1977; Katzell & Dyer, 1977; Dunnette & Borman, 1979; Hunter, Schmidt & Hunter, 1979; Arvey, 1979.)

differential validity are chance occurrences or are indicative of true differences in subgroup test performances. In one such investigation Boehm (1977) examined 31 studies involving 297 comparisons of employment and training selection procedures administered to blacks and whites. Of those 297 comparisons only 8% showed differential validity. A closer look at those studies revealed that reports of differential validity were likely to come from methodological weak studies, i.e., those characterized by small sample sizes, criteria which were not performance measures and/or a weak rationale for hypothesizing that the predictor(s) would have any relationship to the criterion. In sum, she concluded that findings of differential validity were methodological artifacts, and did not reflect any relationships across subgroups. In another study, Katzell and Dyer (1977) reviewed 31 studies and concluded that differential validity did occur at above chance levels in those studies but that differential validity did not favor one subgroup over another (i.e., validity coefficients were lower for whites as often as for blacks). Finally, in an extensive study aggregating 866 validity comparisons of blacks and whites from 39 studies, Hunter, Schmidt and Hunter (1979) identified and controlled for methodological biases. They concluded that findings of differential validity were produced by chance and statistical artifacts, and that therefore differential validity probably did not exist in the population.

If testing practitioners find differential validity when examining their selection procedures but researchers argue that differential validity doesn't really exist, what can we conclude about differential validity and the potential for adverse impact in employment selection? The recent research suggests that careful design of validation studies may prevent differential validity from occurring. Accordingly careful validation, control, and design may eliminate differential validity as a source of adverse impact. However, we must also remember that elimination of differential validity is not a panacea because many other characteristics of the distribution of subsample predictor scores may contribute to test bias and adverse impact. Nonetheless, if carefully designed validation studies are conducted to reduce the possibility of finding differential validity, then these studies will also provide means, standards deviations and score ranges which can be examined to determine their potential for adverse impact.

The difference models which identify unfair use of a test, and provide a rationale and method for setting cutting scores and selecting individuals fairly from differing subgroups, can be used in combination with a good validity study to guide an organization toward more job-related and equitable selection. Models of test bias reflect selection goals of sufficient diversity to accommodate almost any organizational choice about the kinds of selction errors that constitute test bias. One model (Darlington, cited in Peterson & Novick, 1976) even

exists which allows organizations that wish to consider not only performance criteria, but also employment parity or compensatory hiring of minorities in their definitions of test fairness. This model permits explicit inclusion of such factors in setting cutoffs and making selection decisions which are fair in accordance with the organization's values. Regardless of intentions and choice of policy regarding test fairness, an organization's resources may be inadequate for a good validation study because sample sizes (i.e., job openings) are limited. For this reason, cooperative validity generalization studies make the validation process an even more palatable strategy because they can provide data to reduce adverse impact and enable greater work force productivity at reduced costs.

Comparison of Selection Devices

After a variety of techniques have been described and their research reviewed, it is important to step back and put them into perspective. Recommendations can be made, endorsing certain methods for certain purposes. Table 10 is a tool for doing just that. The techniques are compared on the basis of eight important issues for someone choosing a method for personnel selection: (1) reliability, (2) validity, (3) research support, (4) objectivity, (5) face validity, (6) unique data, (7) cost, and (8) skill coverage. Despite the wide variance in the amount and depth of research on each technique, objectivity, cost, skills covered, and psychometric properties, each of the data collection methods is appropriate to use for certain purposes.

The application blank is commonly used by most companies as it is an inexpensive means of gathering background data, including experience and education. Applicants expect to fill out an application blank, thus face validity is good. Additionally, reliability and validity of certain items can be quite high. The disadvantage of the application blank is that it does not directly measure skills. Therefore, where assurance of skill is critical, the application blank would be deficient by itself as a data base for an employment decision. When the employer has little doubt about applicant abilities (e.g., applicant has a license or degree indicating skill level)

TABLE 10

Comparison of Selection Techniques

METHOD	RELIABILITY	VALIDITY	RESEARCH	SCORING CRITERIA	FACE VALIDITY	UNIQUE DATA	COST	SKILL COVERAGE
Interviews	very specific to situation and interviewer; Median Range=.55-.70	very specific to situation & interviewer; Median Range =.20-.25 for JOB PROFICIENCY; somewhat higher for training	extensive, primarily focused on interpersonal perception and decision formulation	subjective judgement of employer	usually good, depending on interview content	provides opportunity to interact with applicant and observe interpersonal dynamics	moderate to high, depending on amount interviewer time, number of interviews and training	intellectual ability, work motivation, interpersonal relations
General Intelligence Tests	Good; Median Range=.80-.90	Higher validity for managerial and clerical occupations. Median Range =.20-.30 for JOB PROFICIENCY; somewhat higher for training	extensive	reliable scoring keys	fair depending on job	provides overall ability measurement	usually low with standardized tests. Increase in costs with administration and scoring time	General Reasoning & Logic; Verbal comprehension & fluency; Numerical comprehension & fluency; Spatial Relations
Specific Skills and Abilitys	Specific to instrument; Median Range=.65-.80	Higher validity for managerial and clerical occupations; Median Range =.25-.35 for JOB PROFICIENCY somewhat higher for training	extensive for most ability tests. Very little on measures of creativity and judgment	reliable scoring keys	usually good depending on job	provides independent measure of skills and aptitudes.	low administration cost; moderate to high equipment costs	Separate measures of: Verbal Ability, Numerical Ability, Spatial Ability, Creativity, Psychomotor Ability

(Table 10, continued)

METHOD	RELIABILITY	VALIDITY	RESEARCH	SCORING CRITERIA	FACE VALIDITY	UNIQUE DATA	COST	SKILL COVERAGE
Personality Tests	Specific to instrument. Highest for Interest tests; Lowest for Projective Tests; Median Range=.60-.70	Higher validity for sales occupations. Median Range=.20-.25 for JOB PROFICIENCY. Usually lower for training	Extensive research in clinical settings. Much less in industrial settings.	General rules provided for scoring; raters may disagree on ratings	poor	provides standardized information on personality characteristics and interests	low to moderate, depending on test administration time and scoring procedures	Interests, Psychological Traits, Psychological Adjustment
Application Blanks	good; median range .87≤r≤.97	median range .40≤r≤.50	quite a bit on items, discrimination, bias	usually objectively coded; depends on how structured questions are	good	not usually, but can ask for goals and self-perceptions (e.g., weaknesses and strengths)	low cost; costs increase with job analysis, keying, continued validating	bio data, job experience, education level
Resumes	not available	not available	none evident	Subjective judgement of employer	good	provides data on style, organization, self-perception	no cost to organization	bio data, job experience, education level, communication ability
Recommendations	not available	standard questions may be valid for some, but not all jobs; when valid .21≤r≤.29*	very limited	subjective judgement of employer	good	data from ex-boss and friends	very low cost (phone call or request letter)	job proficiency interpersonal skill

*based on one study

(Table 10, continued)

METHOD	RELIABILITY	VALIDITY	RESEARCH	SCORING CRITERIA	FACE VALIDITY	UNIQUE DATA	COST	COVERAGE
Work Samples and simulations	generally not available, but design suggests good consistency of application and measurement	median range, verbal work sample: r= .30-.40, job proficiency r=.40-.50, training; motor work sample: r= .40-.50, job r=.30-.40 training	extensive on validity	criteria are explicit; raters tend to agree on ratings	very good proficiency,	job performance shows combination of desire and skill to do job	expensive to design, large amounts time, $ to administer	intellectual ability, interpersonal skills, psychomotor skills (dependir on job sampled)
Thought samples	Not known, but validity coefficients suggest at least .60	high for management entrepreneurial jobs: r=.40-.80	extensive on management and entrepreneurial jobs	scoring categories are well-defined; scorer agreement exceeds .85	good for scoring criteria poor for test instrument	motivational competencies	moderate to high cost: protocols must be hand-scored	interpersonal skills (motive dispositions and ego maturity)
Detectors of Deception	high, but data based on settings not comparable to employment	measured in accuracy terms which are inappropriate for employment application	not related to employment realities	scoring criteria in laboratory settings	poor	data on stress and and attitudes	high to moderate cost to buy equipment and hire trained operator-interpreter	doesen't cover skills

and where cost is an issue, the application blank can be a fairly inexpensive and effective selection tool.*

If an applicant's goals and self-image are important, the resume is a less structured way to access and expand on this information. The cost to the organization is minimal and a resume provides more detail in these areas than does an application blank. Little research evidence is available on the reliability and validity of the resume. Like the application blank, the resume will seldom be sufficient for making selection decisions as skills are not directly measured.

In those cases where interpersonal skills or task related skills are important on the job, recommendations may be a useful addition to the selection process. These are inexpensive and provide a subjective measure of skills which are measured more objectively, but also at greater cost, by other methods. There has, however, been little research on the psychometric properties of recommendations.

The work sample appears to be an especially useful data collection method as it can directly assess intellectual, interpersonal, or psychomotor skills. Psychometric data show work samples to have high validity and good consistency relative to other techniques. The cost may be prohibitive,

---

* Though employers often worry credentials such as licenses and degrees say something about an applicant's competence, Chapter 3 examines this questionable assumption in great detail.

however, as work samples generally require large capital expenditures and may be expensive to administer.

Thought samples provide some of the most useful data available in assessing these institutional components of interpersonal competence. This technique, indeed, is the only way in which motives can be measured. In complex jobs in areas such as management that rely on interpersonal skills, thought samples have shown the greatest consistency in predictive validity over time. They are expensive to score properly, but return a long-term selection benefit far out of proportion to the initial scoring investment.

Detectors of deception such as polygraphs have little to recommend them. Their reliability and validity are tenuous; they tend to antagonize the subject. Nonetheless, in situations where the good will of applicants is not a consideration, where false negative decisions (i.e., identifying an honest response as a lie) are unimportant because of the tremendous cost of false positives (i.e., hiring a dishonest worker), then polygraphs (and such) may provide some unique data of utility for that specific decision situation.

Interviews have high face validity in employment situations and may be particularly useful in assessing interpersonal skills. Though reliability is typically lower than many other measures, with only moderate expenditure it can be raised quite substantially. Validity is low when the interview is used to assess skills other than interpersonal. Therefore, when other

skills are required, the interview will be an inadequate measure. The objective coding of interview data as discussed under Thought Samples provides the best potential solution to problems of reliability and validity. Because of the time required to administer an interview properly, its cost may be prohibitive for its utility as a selection device for many jobs. However, if it is used for other purposes as well, such as public relations, the cost may be more easily justified.

General intelligence tests may be appropriate as measures of overall intellectual ability for lower level clerical or managerial jobs. However, specific tests of intellectual ability show approximately the same cost with higher validity. Cost for both measures will be low if a standardized commercially available test is used.

Specific skill and ability tests are available for assessing intellectual or psychomotor skills. These tests usually appear face valid to the applicant and show good psychometric properties. If tests are constructed by the organization for particular jobs, costs will increase substantially. Tests may also show higher validity for particular jobs if the test uses actual job equipment, but, of course, this will also increase the cost to the organization.

Finally, personality tests show relatively low reliability, predictive validity and face validity but may be especially useful in assessing interests. Also, if psychological adjustment is an issue, these tests may be quite appropriate. The

cost of commercially available tests is usually low, but administration time varies quite widely, and may serve as a source of hidden cost. Scoring keys for particular companies can be constructed for many tests. This will increase the cost but will usually raise the validity of the test as well.

In sum, all the selection techniques discussed, even the most questionable ones, have some advantages. Their unique strengths and weaknesses (in terms of the 8 criteria identified above) should all be considered before making a choice of selection method(s).

III. THE PRACTICE OF PERSONNEL SELECTION

The previous section has described and compared the major employee selection techniques currently in use. The present section is devoted to the application of these techniques in job situations. Nearly all the literature on employee selection deals with the selection instrument as the unit of analysis, rather than with the job or type of job in regard to which a selection decision is made. The intent of shifting attention to the job as the unit of analysis is to examine the appropriateness of a selection system in which multiple sources of data are considered to identify the most appropriate job candidates. Given what is known of the strengths and weaknesses of individual selection techniques, an account of how these techniques are used as the basis for job selection enables the analysis of how organizations carry out matching applicants to jobs.

This section summarizes an empirical study of employee selection practices in the field. The present study was intended to involve a representative sample of jobs to which selection devices are routinely applied, so that the state of the practice in entry level employee assessment could be ascertained for a wide variety of competencies. This study was initiated by defining a sampling strategy based on a comprehensive but simple job taxonomy which would yield to quantitative analysis. Using the taxonomy as a guide, direct contact was initiated with more

than 100 organizations which were expected to represent the most comprehensive employee selection practices, and to elicit from them as much information as they were willing to share regarding their current and past procedures. Following the description of these procedures, this section presents an account of the state of the practice with regard to key issues in employment selection and a statistical analysis of data gathered on 239 jobs that documents current trends in employee selection.

The Scope of the Study

A prime requisite of this research was a job sampling strategy which would be representative of a cross-section of job functions, since it was expected that the majority of formal competency-based selecton systems would be founded on a rational analysis (e.g., Fine & Wiley, 1971) of the tasks and functions performed by individual jobs. The functional taxonomy developed by Katz and Kahn (1978) was particularly appropriate for the present purpose. These authors described their formal taxonomy in terms of five sub-systems which identify formal operations within organizations, and which may also be applied to describe dominant job functions:

1. Production jobs are based on task accomplishment through technical proficiency. These are jobs in which energy is transformed into output and value is added for the organization.

2. Maintenance jobs are oriented toward maintaining stability and predictability within an organization. This may take

the form of preserving existing relationships among other individuals in the organization, or toward the preservation of the status quo.

3. Managerial jobs contain the controlling or decision making aspect of work. People in these jobs coordinate external requirements with internal resources and resolve conflicts among other job functions.

4. Boundary jobs are characterized by their function as linking the parent organization with their counterparts in other organizations. Individuals in these jobs carry out the transactions leading to the procurement and disposal of goods and services.

5. Adaptive jobs are characterized by their roles in organization change, including intelligence gathering, research and development and planning functions. Though these jobs are concerned with innovation, their primary task is helping an organization adjust to a changing environment.

The translation of this functional taxonomy of organizations into a taxonomy of jobs, however, involves adding the level of skill with which a function is required to be performed as an independent dimension. Accordingly, jobs were also classified according to whether applicants were relatively unskilled with regard to the function they would perform in the job; were skilled or moderately proficient in technical and procedural aspects of the job, or were professional in their skill level, meaning that extensive specialized training or experience was acquired prior to selection. Classification of

each job in the taxonomy by skill level was made on the basis of skill level required of the applicant at point of entry into an organization, rather than on the basis of a skill acquired on the job following a suitable period of on-the-job experience or training provided by the hiring organization.

The next task was to identify employers which had documented their selection systems sufficiently well to be helpful in the present study. It was expected that many employers would resist sharing recent documentation of employee selection procedures for reasons bearing on the proprietary nature of selection procedures, the maintenance of confidentiality of the information source, and questions of compliance with EEOC guidelines. To minimize this problem, the first approach taken was to identify specific individuals within target organizations which had conducted selection research and had presented it in both published and unpublished literature, and other individuals who were otherwise known to the authors through personal or professional contact.*

---

* The identification of contacts through the literature review produced mixed results. A significant number of organizations who are pursuing research on their own selection systems could be traced through articles published in academic journals. However, surprisingly little of value regarding the state of the practice was gleaned from a survey of periodicals devoted to personnel practices in specific trades, careers or professional associations. From a random sample of over 100 such periodicals, not a single article within the last five years could be found in which either a rigorous statistical analysis of selection system practices was undertaken or in which data-based research regarding a particular selection device was attempted. Of those articles which dealt with personnel selection practices, the only data available were anecdotal in nature and not useful to the present study.

This procedure obtained access to selection system data on 154 different jobs, from 79 organizations. An additional 85 jobs from 38 organizations were identified through less personal means, including "cold call" contacts with the personnel directors of selected Fortune 500 industrial organizations, and with organizations who had recently advertised employment opportunities in a number of major newspapers, although it was not possible to obtain the same degree of selection system documentation for this additional job sample.

Next, open-ended interviews were conducted with the contact persons within the target organizations. Interviewees were encouraged to provide information about specific jobs for which selection procedures had been documented and to supply that documentation in writing wherever possible. Included in the information sought during this process was a specification of the kinds of knowledge, skills, abilities and other characteristics sought, the types of selection devices that were used to identify them, factors relating to the development and implementation of the selection system, validity and reliability of the system in use, and the number of jobs affected by the process. Data were sought with special emphasis regarding the application of the various sources of selection data discussed earlier, including interviews, objective tests, work samples, simulations and recommendations. As application blanks and resumes are used to gather a qualitatively wider variety of selection data, respondents were asked to indicate whether

biographical data, work history, past performance in similar jobs or education requirements were the important considerations. Additionally, respondents were asked whether or not a license was required for hire in a particular job. No data were available regarding the use of thought samples or detectors of deception in the present sample.

Table 11 provides a summarized classification of the 239 jobs for which selection system data were available. Approximately 25% of the jobs in this table were drawn from "Fortune 500" companies, 30% from other sizable businesses in the private sector, 25% from the federal and state service, education, and other public sector sources, and the remainder from miscellaneous sources. Note that production jobs comprise the largest number of jobs classified according to the taxonomy. It appears reasonable that the majority of jobs that exist are those in which some amount of value is added directly by the employee to the product or service. It is also reasonable that the greatest number of jobs on which selection systems have been documented are production jobs, since value added provides a more accessible validation criterion than do other measures of performance effectiveness. By contrast, the fewest number of selection situations involved jobs classified as adaptive, since most jobs in this category are filled from within the organization through promotion, or from without through highly individualized means of selection which are beyond the scope of the present study.

TABLE 11

Taxonomy of Jobs Surveyed by
Level and Function: A Representative Summary

Skill Level	Job Function: Production	Maintenance	Managerial	Boundary	Adaptive
Unskilled	assembly line worker cashier nonskilled clerical transportation worker waiter stock checker refinery worker miner cook	file clerk employment rep. aide/orderly inspection worker mechanical worker administrative psychiatric attendant	management trainee store manager statistical manager administrative trainee	duty collector checkout computer checkout salesperson human service worker toll collector library clerk revenue officer truck driver	
	n = 32	n = 11	n = 8	n = 18	n = 0
Skilled	medical technician secretary engineering technician food processor mechanic craft worker draftsperson staff nurse field worker	computer programmer administrative assistant firefighter airline maintenance alcohol counselor patrolman hospital corpsman diver	foreman first line supervisor buyer toll facility officer sales manager correctional officer	life insurance agent manufacturer's rep. consultant counselor product service assistant buyer claims authorizer	junior consultant/ trainer
	n = 46	n = 22	n = 22	n = 16	n = 1
Professional	faculty member senior technical specialist pilot architect civil engineer musician nuclear chemist choreographer surgeon	accountant biomedical technician process technician FCR manager	manager of marketing manufacturing manager engineering manager product supervisor plant manager police captain	marketing manager commercial loan officer lawyer technical sales purchasing agent foreign service officer	political generalist management consultant
	n = 31	n = 5	n = 10	n = 12	n = 5
TOTAL	109	38	40	46	6

Trends in Contemporary Practice

Many of these organizations in our sample, particularly those with a long history of selection system documentation provided us with much information about selection practices that would not be reduced to statistical analysis. This section is therefore devoted to some of the more qualitative aspects of employee selection which may serve as an introduction to the qualitative presentation that follows. The concern of this section include the underlying motivation for choosing a given device, the issues and problems that arise with particular selection procedures, and the changes organizations anticipate over the next several years.

First, it should be noted that employing organizations are highly reluctant to share data about their use of employee selection systems. This was not unanticipated, due to the sensitivity and proprietary nature of the information we were selling. However, even after establishing our research credentials through references and correspondence and guaranteeing confidentiality of the data and their sources, nearly 25% of the organization contacted did not wish to contribute to the present study. During our study, we found companies particularly reluctant to share demographic data with us about the applicants. Those who did seem to have data, and were willing to share it, fell into two basic classes: (1) a group of large organizations who were using biographical data explicitly as validated predictors, and therefore were willing to share

applicant characteristics with us; and (2) a group of people who were keeping applicant flow figures, and were in reasonably good positions with the EEOC. Otherwise, companies were particularly defensive and reluctant to give out information about the characteristics of people who went through the process. Their reticence was rooted in a fear that this data could be used against them.

One of the emergent patterns in selection was the use of industrywide selection devices as in the petroleum and insurance industries. In both of these industries, several firms have pooled resources and looked at common needs and selection issues. As part of their cooperative study, these groups examined job analysis, test design and validation. In addition, the petroleum industry has been involved in cooperative studies on validity generalization in hopes of identifying job-related skills and appropriate tests which can be used by the companies within the industry.

When companies described the process of selection in which they were engaged it was revealed that everyone uses interviews. There is, however, great variability as to whether interviews are used as the first screen for an applicant, or whether they are used for a final decision after consideration of other kinds of data such as applications or resumes.

While all organizations reported using interviews, and most organizations reported using skill tests for clerical people, there are other consistencies which emerge in the use of methods

by individual companies within a single industry. For example, biographical data in various formats is used extensively in the insurance industry while public braodcasting companies rely heavily on resumes and trade publications which post job openings. These trends may be reflective of similarities between jobs within a given industry and/or may reflect an industry tradition in terms of selection methodology.

Another trend that emerges in the use of testing as a selection tool is the reduction in the use of personality and intelligence tests by most companies, usually for equal employment considerations. In one case where EEO legislation has had a visible impact on the company's testing process, the firm has dropped testing from its selection procedures, instituting post-hire testing instead. After selection, a person is tested to see where he or she should be placed within the company. Those tests that continue to be used for selection purposes are skill tests which generally have high face validity for the job. These tests are given at lower levels for skilled and unskilled jobs, clerical jobs, and hourly workers.

The extent of structure given to the interview process varies across interviewers, departments and organizations. In our sample, two interviewers are seldom required to use the same structure. While great variability exists between interviews for a single job, variability is even greater across jobs. The finance department might interview people differently from the sales department. Another general finding in the use of

interviews is that more interviews are required for higher level jobs. A professional level applicant will probably be interviewed by many people whereas for lower level positions, the interview process consists of one interview with someone in personnel.

When we asked companies why they had instituted particular selection processes, it was found that the motive behind the selection process was seldom a desire to understand or identify competence for the job. Usually, the selection procedure had been instituted because of a particular problem, such as turnover, or because of equal employment issues. The companies seemed concerned about the need to protect themselves from any suit that might be brought against them, as well as wanting to show affirmative action.

One set of issues of particular interest to us was how the organizations identified the criteria for their selection process, and what these criteria were (i.e., specific knowledge, skills, abilities, or other characteristics). We found two trends along this line, the first being that, in our sample, rigorous job analysis was clearly the exception rather than the rule. Very few companies had actually conducted job analyses. The second trend was that the domain of job requirements tended to be identified by talking to the hiring supervisor and generating a list of characteristics the supervisor felt to be necessary for the open job. One reason infrequent use of formal job analysis could be that organizations tend to think of job

analysis as appropriate for the development of paper and pencil tests, but not as an important step in developing a method to evaluate resumes, application blanks, or as a prerequisite to the interview. Since these latter techniques are more often used than paper and pencil tests, job analysis is less common. The sample was asked how the selection process was originally developed. The practitioners, however, seldom had knowledge of the process development. Typically the process was described as evolving over a long period of time or else it was empirical, not based on a concept of what was important to the job, but rather, on finding items which would empirically predict performance. Organizations with empirically keyed selection tests (such as scored biographical data forms) must continually revalidate them and therefore are more in touch with the development of the process. Trait or skill measures are more likely to be considered logically related to job performance and therefore receive less questioning of the rationale for their use.

The organizations mentioned three other ways their selection procedures were developed. By far the most typical was using published tests that already had established norms as well as documented reliability and validity data for similar jobs. In addition some companies spoke of developing their own inhouse tests. The fewest number of organizations mentioned hiring outside consultants to develop some of the more sophisticated kinds of selection techniques, such as assessment centers for selection or coded interviews.

When organizations were queried about the validity of their selection process, it was revealed that few organizations had conducted actual validity studies. Those that had done validation primarily examined the validity of psychometric tests, especially those used for selection into high-level positions. Also tests were more likely to be validated by the company if their relatedness to the job wasn't immediately clear. Personality tests with their low face validity frequently fell into this category. For lower level positions, tests in use were seldom validated. Instead the organizations relied on published validity statistics.

Interviews were also very seldom validated. Only a few companies were concerned with the validity of their interviews because they didn't think of them as selection tests. The only other selection device which has been examined for validity is biographical data. These data were usually validated for particular jobs through keyed application blanks. Other methods, such as recommendations, standard non-keyed application blanks and resumes, tend not to be validated forms for collecting biodata.

When representatives of organizations were asked how their selection process had worked, they relied on anecdotes, face or content validity, as evidence that their ongoing selection process was selecting the right people. They did not rely on more empirically rigorous methods for demonstrating the utility of their selection process. Where there was empirical evidence,

that evidence tended to be gathered for the more complex, or less face valid procedures such as biographical data, personality tests, or assessment centers.

A few organizations reported that the problems they had encountered in the use of various selection processes were that an instrument was not seen as being face valid, or that the selection procedure was not accepted within the company itself, usually because of time or money constraints.

One large trade organization for the insurance industry also stressed that biographical data need to be revalidated over time, since such data could not be counted on to maintain their validity over different years and different applicant populations. This is a problem which is grossly overlooked by organizations using other kinds of selection methods. The biodata's instability over time is a severe problem for that kind of data. We found few organizations outside of the insurance and petroleum industries which revalidated their selection procedures.

Organizations involved in validity research stress misuse of selection procedures as a problem. They found it difficult to conduct adequate research because operating groups often selected individuals on the basis of an unvalidated pilot instrument. Cost was also a hindrance for some sophisticated procedures such as assessment centers. This was particularly true when large capital outlays were necessary, and techniques were unfamiliar or nontraditional.

Our sample reported that data from selection procedures were usually maintained in company personnel files. Some companies

kept data in regional or central files with the intention of doing later research. Many organizations were sceptical about inquiries around the use and maintenance of selection data files. We can only assume that this scepticism arose from EEO concern. Most organizations reported no further use of selection data but those that did responded that the data were most often kept for one of three purposes: for maintaining compliance records (EEOC and Affirmative Action), for giving feedback to the applicants, or for validation on the instrument or selection techniques.

Organizations were asked what changes they anticipated in their selection process. Several companies predicted a trend toward increased job analysis, increased validation efforts (especially for such devices as the interview), and more behavioral methods of selection. Additionally, one organization spoke of an anticipated need to justify selection and promotion decisions to the applicant. In the past this had not been required but the anticipation seems reasonable.

When asked whether educational credentials were used as part of the selection process, those organizations that did use or consider educational credentials, did so primarily for high-skilled jobs. Organizations seldom specified either the validity of the educational credential as a predictor of job competence, or the weight given to the information about education when making a selection decision.

Analysis of Trends and Practices

The following data were noted for each of the 239 jobs, outlined in Table 11, to permit a systematic quantitative analysis: the taxonomic job classification (job function and level of skill required), the sources of data considered in making the selection decision, changes in use of each data source over time (e.g., whether the data source increased, remained constant, or decreased in use), and whether the data source had been validated with respect to the job by the hiring organization.

Based on supplementary information and notes taken during the interviews the generic competencies required by the job were also recorded. These competencies fell into three broad groups: (1) intellectual competencies, including technical knowledge problem solving, planning and organizing, and abstract reasoning; (2) interpersonal competencies, including communication, leadership, counseling, teaching skill and self-presentation; and (3) psychomotor competencies, including manual dexterity, agility, physical strength, eye-hand coordination, and speed and accuracy. Each job was coded by whether it required one or more of the competency group rather than whether an employing organization selected for these competencies.

Table 12 presents a summary of the descriptive statistics regarding the use and validation of eleven sources of selection data. The interview was, by far, the most frequently used

TABLE 12

Use and Validation of Selection Data

	Source of Data	Frequency of use	Validated use	As percentage of frequency of use: (Decreased Use / Constant Use / Increased Use)		
1.	Interview	73%[1]	7%	(0%	97%	3%)
2.	Skill and Ability Test	38%	35%	(5%	92%	3%)
3.	Work History	38%	3%	(0%	100%	0%)
4.	Education	31%	5%	(18%	82%	0%)
5.	Biographical Data	21%	33%	(8%	89%	3%)
6.	Recommendation	21%	0%	(0%	100%	0%)
7.	Personality Test	18%	69%	(15%	83%	2%)
8.	Past Performance	14%	6%	(0%	100%	0%)
9.	Work Sample	12%	17%	(0%	94%	6%)
10.	License	8%	0%	(0%	100%	0%)
11.	Simulation	5%	45%	(8%	84%	8%)

[1] Probability underestimated: see text

source of selection information. The figure of 73% is in all likelihood a low estimate of the number of times both formal and informal interviews are conducted to screen employees, given Scott's (1961) figure of 98% in his study of interviewing practices and the observation that many of the respondents in our sample did not volunteer information about use of the cursory personal interview. Data from skill and ability tests and from work histories, are therefore probably considered less than half as often as the interview. Despite the interview's popularity, or perhaps because of it, the interview was among the least well-validated sources of information used in making a selection decision. Among the sources of selection data listed, work history, educational background and recommendations were also frequently used yet seldom validated. The most widely validated source of information was the personality test, due mainly to the fact that most employers relied on published validation statistics to justify their use. Skill and ability tests and biographical data were two additional sources of information that were frequently used and relatively well validated in the field. The performance simulation, largely because of its complexity and the expense involved in its administration, received significant field validation, despite its infrequent use.

The remaining data in Table 12 are devoted to trends in the use of the various data sources. As a rule, the use of data sources has remained substantially unchanged for each selection

system. Nevertheless, conclusions may be drawn from the proposition of instances in which the use of a particular data source was increased or decreased over the life of a selection program. Specifically, the educational requirement and the personality test show trends toward decreasing use, undoubtedly reflecting recent court decisions striking down educational requirements as unnecessarily discriminatory and personality testing as not demonstrably relevant to job requirements. In the absence of trends toward the increased use of other data sources, one can conclude (1) that the selection interview, due to its mere pervasiveness, is being counted upon more and more heavily in making the final selection decision, even though it is among the least reliable and valid of the selection techniques; and (2) that the use of credentials and personality test, though viewed as important sources of selection data, is simply being discontinued for fear of reprisal in the courts.

Somewhat surprisingly, few differences in the use of particular data sources were found between major job functions. The data in Table 12, therefore, present a largely accurate picture of the use of data sources for production, maintenance, managerial and boundary jobs. A few differences by job function, however, were obtained at statistical levels of significance ($p < .01$). Skill and ability testing was nearly twice as prevalent among maintenance jobs than among others, and education and simulation data were considered twice as often for managerial jobs than for production, maintenance or boundary jobs.

Table 13 illustrates the characteristics of the job sample in terms of the competencies required by the job according to job function and skill level.* In the analysis of competencies by job function, the majority of managerial and adaptive jobs in the sample require some specified level of intellectual competence; the majority of managerial, boundary, and adaptive jobs were found to require interpersonal competence; and production and maintenance jobs largely require psychomotor competence. The second part of Table 13 shows the data collapsed across job functions and examines required competency broken down by the skill level specified by the job requirement. The great majority of unskilled jobs require psychomotor competencies to the relative exclusion of intellectual-interpersonal competencies, while jobs designated as skilled lean toward the requirement of psychomotor competencies but are, in the main, more balanced in their requirements of the three generic competency groups. The professional skill level jobs in the sample require a much higher degree of intellectual competencies than interpersonal or psychomotor competencies, although these last two competency groups are significantly represented. Considering all jobs in the sample together, the

---

* Note that the percentages listed which are percentages of the row total exceed 100%, since many jobs are represented under more than one competency.

TABLE 13

Characteristics of Job Sample: Breakdown by Job Type and Skill Level by Competencies Required by the Job[1]

Job Function	n	Competency Required by Job: Intellectual	Interpersonal	Psychomotor
Production	(109)	36%	9%	92%
Maintenance	(38)	34%	24%	68%
Managerial	(40)	65%	85%	0%
Boundary	(46)	28%	78%	3%
Adaptive	(6)	100%	83%	17%
All Jobs	(239)	41%	39%	59%

Skill Level	n	Competency Required by Job: Intellectual	Interpersonal	Psychomotor
Unskilled	(69)	10%	26%	74%
Skilled	(107)	34%	46%	58%
Professional	(63)	86%	43%	44%
All Jobs	(239)	41%	39%	59%

[1]Percentages are based on proportion of row totals

competency representation, although favoring psychomotor competencies to some degree because of the large number of production jobs surveyed, appears reasonably balanced.

With these descriptive statistics as background, trends in the use of particular selection data were examined by the level of skill and the competencies required by each job. Table 14 presents these data for the eleven data sources of the previous table. A consistent finding is that the frequency of use of selection data varies directly with the level of skill required in the job. This finding is consistent for the interview, work history, education, recommendation, past performance and license. Skill and ability tests show the only significant departure from this trend: employers who seek professional level skills tend to shun objective testing in general as a method for ascertaining competency, while significant numbers of employers who are hiring unskilled and skilled employees may not be able to assume the presence of minimal skills or abilities in applicants and therefore find skill and ability testing to be useful additions to a selection system. However, with all sources of selection data taken together, the median number of data sources considered across jobs is 1.3 for unskilled jobs, 1.8 for skilled jobs, and 3.3 for professional jobs, a highly significant linear trend in the use of selection devices with particular emphasis in professional-level jobs.

The final tables in this section illustrate the likelihood with which sources of selection data are used for jobs requiring intellectual, interpersonal or psychomotor competencies. As

TABLE 14

Frequency of Selection Data Use Related to
Level of Skill Required in Job

Source of Data	*LEVEL OF SKILL* Unskilled	Skilled	Professional	Significant Trend
1. Interview	65%	70%	86%	linear ($p < .01$)
2. Skill and Ability Test	36%	50%	13%	nonlinear ($p < .001$)
3. Work History	22%	32%	67%	linear ($p < .001$)
4. Education	20%	22%	49%	linear ($p < .001$)
5. Biographical Data	28%	20%	18%	
6. Recommendation	7%	21%	36%	linear ($p < .001$)
7. Personality Test	17%	17%	14%	
8. Past Performance	6%	8%	35%	linear ($p < .001$)
9. Work Sample	10%	10%	16%	
10. License	3%	6%	19%	linear ($p < .001$)
11. Simulation	4%	5%	5%	
Number of Cases:	69	107	63	

the use of selection devices was shown to be related to skill level, which was itself correlated with type of competency required by the job (see Table 13), the figures in these tables were adjusted for the correlation between level of skill and competency requirements. Table 15 shows the relative degree to which a source of data will be employed as a function of whether or not a particular competency is required by the job. It is evident here that work history, education and simulations are more likely to be considered when intellectual skills are required than when they are not; skill and ability tests are the only devices that are by themselves more likely to be used when interpersonal skills are required; and skill and ability tests and licenses are more likely to be considered when psychomotor competencies are at issue.

In most cases however, it is likely that more than one source of selection data will be used, so it is necessary to examine which data sources tend to be used together most often in selecting for certain competencies. To this end a discriminant analysis was performed for each of the 11 sources of data considered by whether or not a specific competency was required for the job, and Table 16 presents the results. Taken together, education and simulation data are more likely than chance to be considered in selecting for intellectual skills while, curiously, skill and ability tests are ignored to a degree greater than would be expected by chance. Simulations, personality tests and recommendations were found to be used more often when interpersonal competencies were required by the job, while skill

TABLE 15

Frequency of Selection Data Use Related
To Competency Requirement of the Job[1]

	Is Competency Required?	Intellectual No/Yes	Interpersonal No/Yes	Psychomotor No/Yes
1.	Interview	71%/77%	76%/71%	73%/73%
2.	Skill and Ability Test	48%/20%**	25%/43%**	28%/42%*
3.	Work History	30%/53%	37%/40%	39%/38%
4.	Education	17%/46%**	29%/28%	40%/21%**
5.	Biographical Data	20%/22%	22%/21%	27%/17%
6.	Recommendation	17%/29%	17%/27%	22%/21%
7.	Personality Test	17%/15%	13%/22%	25%/10%**
8.	Past Performance	12%/20%	12%/18%	17%/13%
9.	Work Sample	11%/14%	13%/10%	8%/15%
10.	License	8%/10%	12%/3%*	2%/13%**
11.	Simulation	1%/10%**	1%/10%**	10%/1%**

*p < .05
**p < .01

[1]corrected for level of skill required

TABLE 16

Selection Data Considered as a Function of Competency Required by the Job

Competency Required	Data Sources Most Likely to Be Used	r[1]	Data Sources Least Likely to Be Used	r
Intellectual	Education	.27**	Skill and Ability Test	-.23*
	Simulation	.17*		
Interpersonal	Simulation	.19**	Skill and Ability Test	-.20**
	Personality Test	.11	License	-.12
	Recommendation	.15*		
Psychomotor	License	.19**	Simulation	-.23**
	Skill and Ability Test	.14*	Personality Test	-.20**
	Work Sample	.11	Education	-.21**

[1]Correlations are with requirement of skill (0 = not required by job/ 1 = required by job), corrected for skill level of applicant.

*p < .05

**p < .01

and ability tests and licenses were considered less often. Finally, when psychomotor competencies are required, licenses, skill and ability tests and work samples tend to be used individually or in combination in making a selection decision, while simulations, personality tests and educational background are the least likely sources of data to be considered. It is interesting to note that many of the data sources which are likely to be used in selecting for intellectual and interpersonal competencies are the least likely to be used in selecting for psychomotor competencies, while those sources of data used in selecting for psychomotor competencies are correspondingly less likely to be used in selecting for either intellectual or interpersonal competencies.

Implications for Educators and Employers

It seems reasonable to expect that the selection devices used by an employer to make a hiring decision would reflect in some way the job competencies required of the applicant. In practice, the degree to which this is true depends on how broadly the job competencies are defined. If one is concerned with selecting for specific skills, abilities or performer characteristics such as manual dexterity, negotiating skill, the ability to delegate tasks, knowledge of the latest health care technology, or planning skill, the state of the practice in selection is woefully inadequate. The primary reason for

this is not the often low reliability or validity of individual selection devices, but rather the general absence of job analyses on the basis of which selection devices could be developed. Selection procedures are chosen without much evidence that they tap a relevant domain of job-related skills. With the exception of training programs with practitioner-teachers (for example, in the health care professions), there is very little contact between the world of education and the world of work. That will help educators do a better job preparing students for life outside the classroom.

In reality the selection procedures used by most employing organizations provide little insight into the specific knowledges, skills, abilities and characteristics required for work. Without job analyses to inform selection systems, it is surprising that any validity at all can be obtained by most selection procedures currently in use. If measures of competency are undertaken at a more general level for selection, however, some reassuring observations can be made. The present empirical analysis of the state of the practice was focused at generic competencies--intellectual, interpersonal and psychomotor--then specific competencies within each of these categories. With face validity being a prerequisite for the use of selection techniques and the nature of inferences drawn from the data collected through the use of these techniques, current selection practices are more likely to detect competencies at the generic level which are related to job

requirements. For example, most jobs classified as boundary positions emphasize interpersonal competencies (see Table 13), and the selection interview, a device which is sensitive to general interpersonal skill, is used in the majority of employment situations requiring interpersonal skill (disregarding for the moment that the ubiquitous interview technique is used to select for nearly all jobs). Therefore, applicants for jobs which serve a boundary function have a high chance of being assessed on at least one relevant job dimension (interpersonal skill) given current selection practices, even in the absence of a thorough job analysis. Table 17 illustrates specific competencies grouped by generic competency category and measured by the sources of data considered in the previous sections of this chapter.

However, simply because a selection device is particularly sensitive to certain generic competency dimensions, that does not necessarily imply that information related to relevant dimensions will be extracted by the employer in making a final selection decision. One can only estimate the maximum availability of relevant selection data by comparing generic competencies required for a job and the use of a particular selection procedure which has a disposition to focus on these competencies. The data collected on the 239 jobs allowed such a comparison to be made. The generic competency measured by each selection device (See Table 17) was compared with the

TABLE 17

Generic Classification of Typical Competencies Obtained Through Employee Selection Data Sources

	Variables Tapped in Each Generic Competency Area		
Generic Competency:	Intellectual	Interpersonal	Psychomotor
Data Source:			
Interview	General intelligence	Sociability Interactions with others Communication ability	
Skill and Ability Tests	Verbal comprehension Verbal ability Numerical comprehension Numerical ability General reasoning Creativity and judgment Spatial realtions		Manual dexterity Visual acuity Auditory acuity Finger dexerity & speed
Personality Tests	Intellectual dysfunction	Sociability Dominance Sensitivity Independence Extraversion Cooperativeness	
Biographical Data $_{ra}$	Educational level Degree status Grades (GPA) Interests	Social network	
Experience $_{ra}$	Job proficiency	Job proficiency	Job proficiency
Educational Level $_{ra}$	Area of specialization Years of study Grades Academic honors		

(Table 17, continued)

	Variables in Each Generic Competency Area		
Generic Competency:	Intellectual	Interpersonal	Psychomotor
Data Source:			
License$_{ra}$	Content mastery Recognition of principles Application of knowledge to problem solving		Ability to manipulate equipment for accomplishment of task
Recommendation	Job proficiency	Responsibility Cooperation Independence Sociability	Job proficiency
Past Performance	Job proficiency	Job proficiency	Job proficiency
Work Samples/Simulations	Planning and organizing Decision making Creativity Initiative Judgment Thoroughness Ability to learn Analytical skills	Communication Leadership Delegation Empathy Listening skill Teamwork Human relations Teaching and supervising	Coordination Strength Steadiness Positioning Seeing Hearing Balance Speed

r = accessed through resume
a = accessed through application

required generic competencies of each job to derive an indication of how likely selection systems were to produce data on required intellectual, interpersonal or psychomotor competencies (whether or not these data were ultimately used in making a selection decision). For 95% of the cases in which either intellectual or interpersonal skills are required by the job, selection procedures were used which were appropriate sources of data on these competencies. In only 80% of the cases where psychomotor skills were required, however, did the selection procedures used have the potential to provide useable data in this competency area, due primarily to the fact that the interview, an inappropriate source of data on psychomotor competencies, is used in many of these jobs in preference to more direct measures of psychomotor skills.

This finding raises the issue of using selection devices which have a high probability of yielding data that are inappropriate to the requirements of the job. In reference to the previous example involving selection for psychomotor competencies, it is clear that reliance on the interview as the sole source of selection information may tempt an employer to base a selection for a job requiring psychomotor skills on information related to intellectual or interpersonal competencies which have either a tangential or a non-existent relationship to the job requirements. Indeed, in over 70% of the cases in which psychomotor skills were not required by the job, selection devices were used which had at least the potential of providing

information on such skills. Similarly, in 90% of the cases where intellectual competencies and interpersonal competencies were not required, selection devices were employed which provide data on these competencies. These percentages are probably inflated as representations of the actual use of inappropriate data in making a selection decision, but nevertheless they illustrate present dangers of such decision-making processes. As it is seldom clear even to employers what competencies are required by a job, further errors introduced by the application of inappropriate selection criteria make it difficult to predict whether an otherwise competent individual will pass the selection hurdle.

In summary, the major findings of the present empirical study of data sources considered in making a selection decision are as follows:

(1) With the exception of the personality test, the majority of selection devices and data sources have not been widely validated by employing organizations that use them. The selection interview in particular, though it is used by over 90% of employers, is not only among the least reliable and valid selection devices in existence but is also among the least validated of data sources within employing organizations.

(2) Differences in the use of selection devices do not vary with job functions themselves, but vary with the type of competency required by the job and with the level of skill

expected of the employee at the point of selection. The emphasis on the use of different selection data sources, as a general rule, increases with the level of skill required in the job, independent of the types of competencies required.

(3) General trends in the decreasing use of education and personality testing for employment selections, reflecting court decisions on cases involving employment practices and problems arising from attempts to comply with EEOC guidelines are not balanced either by trends in selection system validation or the increasing use of other selection devices. This suggests both an increasing reliance on other, perhaps less reliable or valid information sources and a lessened emphasis on entry selections in general.

(4) Though in many instances the techniques which are used to select for particular competencies are appropriate to the task, there are significant numbers of instances in which inappropriate sources of data are used to select for certain competencies while other, more appropriate measures are ignored. This appears to be due to both a general lack of good job analysis among the jobs we surveyed and the overuse of data services which are interpreted beyond their limitations.

IV. CONCLUSIONS

Discrepancy Between Ideal and Actual Practice

Throughout the preceding pages we have described, analyzed, and criticized both the research and the practice in competence assessment for selection. The practice of selection has been critiqued by comparison with the research on selection methods and by comparison with more generalized prescriptions for improving employment selection. Research on a variety of methods has likewise been compared favorably with the ultimate goals of near-perfect reliability and high validity. While the individual criticisms and comparisons throughout the text stand on their own, this is an appropriate point to attempt to tie them together and make a statement about the discrepancies between ideal selection through systematic competence assessment, and the reality of selection practices in functioning organizations.

As early as 1913, Munsterberg was studying street car motormen and prescribing methods for improving selection for that job. At about that same time Otis was doing the basic research on paper and pencil selection tests which he later turned over to the Army for the development of the famous "Alpha" and "Beta" tests used to select World War I soldiers. Obviously, selection testing is not something new. For more than 65 years employers been trying to develop devices to identify people who will be competent workers in the organization. Since the very early days of selection testing,

certain practices have been prescribed for optimizing the validity and utility of selection techniques. It is both heartening (in terms of foresight) and distressing to notice how appropriate those prescriptions are to this day.

In 1923 Freyd published a series of three journal articles in which he outlined in detail the principles and practices of "Measurement in Vocational Selection." He argued that an organization should identify that department in which the greatest savings could be realized from improved selection. The actual design of a selection instrument should then follow 10 steps.

1. Do a job analysis to identify what leads to success or failure on the job.
2. Identify a single measure of the criterion of success.
3. Select a sample. Inexperienced applicants are preferred to incumbents. Identify and study age or sex differences.
4. Develop an exhaustive list of abilities required for the job and recommend procedures for evaluating each one.
5. Find or devise appropriate measuring instruments (not restricted to tests).
6. Administer the tests and measures under carefully controlled conditions.
7. Statistically compare testing results with the criterion.
8. Combine multiple measures for maximum correlation.

9. Justify the measuring instruments by comparing their predictive accuracy with selection methods already in use.
10. Only if the comparison is favorable should the new measures be installed, being sure they are properly used. Continuously monitor their predictive accuracy and adjust them to the type of applicant and changes in industrial demands. (Freyd (1923) as cited in Guion, 1965.)

As noted by Guion (1976) these "tenets of orthodoxy" are remarkably contemporary. Throughout this chapter we have made the exact same recommendations. We have criticized many current methods of selection because they don't specifically identify and tap job skills derived from job analyses (1). The criterion for employee selection is often unclear (2). Where criterion related validation is attempted, it is often of the concurrent variety, i.e., incumbents are measured against their current performance (3). Selection techniques often tap only one skill or a small subsample of the skills needed for the job (4,5). Selection procedures such as the interview are not consistently applied to all candidates (6). Statistical validation is avoided in favor of conceptual content validation (7). Decisions are made on the basis of a single piece of selection data such as one interview or an aptitude test (8). Cost benefit analyses are foregone (9) and new methods are quickly implemented (10). Freyd's caveats have gone unheeded to a large extent.

While not all selection instruments have been as rigorously validated in research as Freyd would hope, academic research in this area exceeds operational research in volume and quality. Job analysis, rigorous test design and criterion related validation, however, have remained relatively uncommon in practice. This discrepancy between the practice and the research is discouraging. The methodology for doing all three of these things is well documented. All that is really needed is time and money to implement methods already proven by research. The impact of new research on selection practices would be marginal compared with that of the employment practice catching up with the current state of the art.

In his 1976 review of selection practices for the Handbook of Industrial/Organizational Psychology (Ed. Dunnette), Guion contends that the aforementioned tenets of orthodoxy are still contemporary as guides to good selection practice but that they are often overlooked. He concludes that the 1970 EEOC guidelines and court rulings in support of validation would bring more companies back to the rigorous selection practices prescribed in 1923 and today. While this seems to have been a logical prediction, and clearly the intent of the EEOC and Congress, our current data do not bear witness to such a trend.

Factors Underlying the Discrepancy

Considering the ever-growing body of research in employee selection technology, why has there been so little progress in

the state of the practice? Though typical applications of the selection procedures discussed yield poor reliability and validity, carefully managed studies involving job analyses and performance criterion validation have shown that many of these procedures could be of high utility in employee selection. From a competency based perspective, predictive validation is the key to improved selection processes; that is, the data obtained through a selection procedure should be related in some way to the competencies required for performance of the job. A demonstrated criterion relationship is important for choosing a selection process that has utility in prediction (over chance levels of accuracy) and is cost-effective. Why, then, are there so few validation studies undertaken to demonstrate these relationships?

The use of selection procedures satisfies an organization's need to make a "careful decision". In their own eyes selection data must first have a face-valid relationship with performer qualities that seem to be important on the job. In most cases, this satisfies the organization; studies are seldom undertaken to test the criterion-related validity of face-valid tests that reflect employers' assumtions about job-related characteristics. In addition, the uniform guidelines, despite their intention to apply fair and consistent standards for the evaluation and use of selection devices, have made selection testing less appealing to employers who feel that data used in making a hiring decision can be used against them in a discrimination suit. Also affecting the choice of whether or

not to validate is that most employing organizations are simply too small in terms of jobs and people to make test validation feasible. Either the costs of such a process are prohibitive or there are too few people in a given job to make any validation statistics meaningful. Since many organizations either ignore the state of the art and remain satisfied with face validity, see the EEOC guidelines as a threat, or are simply too small to afford validation, it is a wonder that validity studies are carried out at all.

All these trends, however, are counter to the intuitive notion that good, valid selection should lead to greater organizational performance and productivity. Why organizations find the trade-offs too great in pursuing this goal seems a mystery. Theirs is, in essence, a reactive mode which suggests why the state of the practice has not improved over the past 50 years. Despite these factors, a number of organizations in our sample have persisted in attempting to validate their selection precedures. Data on those organizations suggest a model for explaining differences in the nature of hiring organizations themselves which, in combination with internal and external pressures on the organization, largely determine whether or not validation of selection procedures will be pursued. Likewise, it will be proposed here that the lifecycle of an organization has a strong impact on the state of the selection and validation strategies used in that organization.

Wright (1969), Rothschild (1976), and others have related business planning to certain stages of organization

maturity and strategy. These stages largely characterize the behavior of an organization with respect to the role of management, strategic planning, policies, procedures, and information and measurement systems. Organizations may be characterized by four stages of maturity:

1. The embryonic organization is characterized by small size and initial rapid growth, technological change, and pursuit of customers for its products and services. Leadership in such an organization tends to take an entrepreneurial, seat-of-the-pants approach, and such an organization is therefore highly dependent on the environment for its continued support.

2. A growth organization is still growing rapidly but tends to be larger, better known and established and in a product or service area where new entry by other organizations is more difficult. Management, like the organization, remains market-oriented and is often torn between fulfilling boundary and managerial roles.

3. A mature organization is still growing albeit at a less rapid rate; holding and defending the products and services it has established is a primary concern. Here the managerial role becomes more administrative and the organization adopts a goal of stability and maintenance of market share.

4. An aging organzation is faced with a lessening of demand, competition, and diversity of products and services. Management becomes opportunistic and relatively short-sighted. The goals of the organization are either to harvest the product for all it is worth, or to divest.

The proposition here is that the strategy to validate the use of employee selection instruments is bound together with the emerging organizational character defined by its level of maturity. A second factor affecting the likelihood that an organization will or will not undertake selection validation involves sophistication about testing and selection dthroughout the corporation and in the person of the personnel administrators who choose and use selection methods. Sophistication in the use and design of selection instruments is likely to increase as an organization passes from the embryonic to the growth and then to the mature stage. By sophistication we mean the insight gained from experience with successful and unsuccessful selection attempts, as well as formal knowledge of the difficulties involved in defining selection criteria, constructing predictors, analyzing predictions, and refining the use of the selection device for making equitable job-related employment decisions. Since the personnel function evolves and is elaborated over time as an organization grows (Katz & Kahn, (1978),sophistication in use of selection procedures should grow as the personnel function grows and the personnel staff becomes more specialized.

If you consider for a minute the orientations of organizations at different stages, it is clear that personnel functions have an increasing potential over time to help an organization maintain a stable, effective work force as the organization approaches higher levels of maturity. When an

organization is aging, however, the personnel function may atrophy because there is a decreased need for having, and no long-term commitment to having, the best work force in the face of less competition or decreased market share. Nonetheless, regardless of the stage an organization has reached in its lifecycle, more sophisticated corporate decision makers and personnel staff will be informed and presumably more rigorous in their efforts at design and evaluation of selection procedures.

Factors outside the organization also come into play in determining what strategy will be used for selection. One factor that stands out as an important determinant of the ultimate selection strategy is whether or not an organization is monitored by the Federal government or the courts with regard to its past and present selection practices. Monitoring may come about through past violations of Title VII which have surfaced in the courts, or by virtue of the organization's being heavily regulated by government agencies.

Our observations suggest that regardless of the stage of the organization's maturity and selection and testing sophistication, if the organization is vulnerable to being monitored, the strategy it will adopt is to document its selection practices rather than to validate its procedures or to continue with undocumented past practices. If one accepts that the first goal of an organization is survival and the second goal is growth, then it is unlikely that public scrutiny

of the organization's selection and hiring practices will be countered by amassing voluminous statistics documenting that no adverse impact occurs (i.e., hiring to conform to the 4/5 rule of the 1978 uniform guidelines on selections), but which do not necessarily document the validity of the selection criteria. This trail of paper is essentially a protective measure. Even if validation is to be undertaken ultimately, documentation of present practice, applicant flow, and impact is a necessary protective first step.

Figure 1 shows the logical relationships between stages of organization maturity, external factors, selection and testing sophistication, and the ultimate selection strategy adopted by the organization.

In light of our sample survey and of the factors we have identified above, this figure presents (in the form of Venn diagrams) our basic hypotheses about (1) why and when an organization devotes to a selection (hiring) process is a function of three sets of variables:

1. Stage of organizational development (the life cycle);
2. Visibility of the organization to previously excluded groups, consumers, monitoring agencies; and
3. Personnel and corporate sophistication about testing and selection.

At a very general level of analysis which will allow us to compare corporate resources of various types on one scale, we have identified four different levels of resource allocation to

Figure 1

Combinations of Factors Determining
Resources Devoted to Organizational Selection Practices

The interaction of circular areas depicting stage and the other factors represents the resources devoted to selection.

In each case the stage of organizational development of life cycle is given.

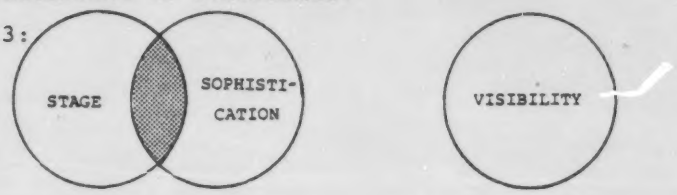
CASE 1:

When an organization is neither visible nor sophisticated in selection practices there is no intersection and no extraordinary resources are devoted to selection.
SELECTION IS AD HOC.

CASE 2:

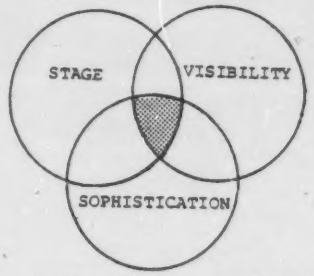
When an organization is visible but not sophisticated in selection practices resources will be put into the short range protective stategy of record keeping.
SELECTION IS DOCUMENTED.

CASE 3:



When an organization is sophisticated in selection practices, but not visible, resources will be put into efforts to improve the accuracy of predictions of methods.
SELECTION IS VALIDATED.

CASE 4:



When an organizaion is both visible and sophisticated it will be under short term pressure to protect itself and long term pressure to improve selection practices.
SELECTION IS DOCUMENTED AND VALIDATED.
Some resource expenditures will benefit both efforts but overall resources required will increase.

selection practices. They include: (a) Ad hoc selection which requires no extraordinary resources to function; (b) Documentation which requires time, money and manpower to record and maintain demographic data on applicant flow; (c) Validation which requires time, money, manpower, and a research and development commitment; (d) Documentation and Validation which requires a greater resource pool to accomplish both. Figure 1 suggests which level of resource allocation will follow from a given combination of the three sets of factors we have identified.

In addition to using the three sets of factors to identify what level of resource expenditure can be expected in a given firm, we can hypothesize how visibility and sophistication will be distributed across stages of development. Given that distribution we can also predict during which stages the goal of validating selection measures of competence is likely to be attained. Thus, on the basis of what we observed in our sample survey we propose:

1. Visability is low for new, young, embryonic organizations which have yet to establish themselves in the public eye or grow to proportions qualifying them for monitoring. Otherwise, visibility is constant with a slight decline possible among aging firms that have survived previous monitoring.

2. Sophistication increases as an organization develops (see Katz & Kahn, 1964), evolving a personnel function and

expanding its influence. This function and its forces atrophy in aging organizations with decreased hiring and less active competition.

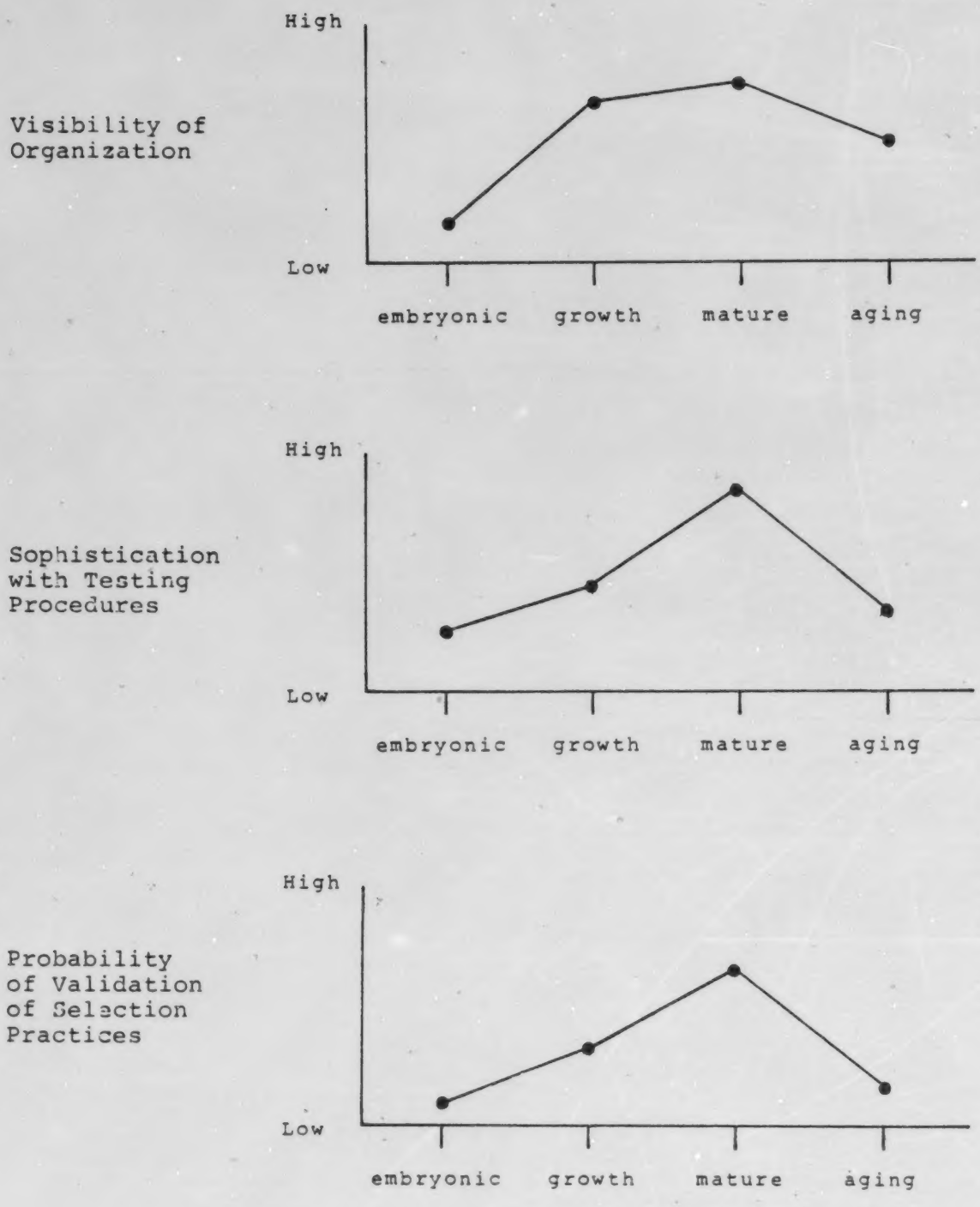
3. Validation (or the documentation/validation combination) is likely to occur in mature organizations most often, followed by growth organizations, and is least likely in early or late stages of an organization's life cycle.

Figure 2 reflects the relative likelihoods of these conditions. The axes on the probability distributions are not scaled because at this point the graphs are only suggestive. These predictions reflect findings from the present state of the practice survey. In general, industries on the decline (for example, supermarkets and railroads companies) reduce their use of formal assessment procedures, while organizations which are both on the increase and have a vested interest in maintaining future stability (for example oil and other energy-related industries, and insurance companies) show an increase in selection research. But even when an organization is mature, growing, and seeking stability, it still must be relatively free of external scrutiny or must also possess a management team with a sophistication in selection procedures which can produce a long-range manpower strategy to support a validation effort. Otherwise the tendency will be either to document past practices or to continue selecting on an ad hoc basis.

This model raises interesting implications for a prescription to change the state of the practice in employee selection.

FIGURE 2

Relationships Between Organizational Maturity and Factors Related to Selection Validation



As was originally envisioned, the EEOC guidelines were intended to have a facilitating effect on ensuring validation of selection procedures while maintaining fairness in hiring. Recent court decisions, however, have resulted in employers reducing their reliance on objective testing for selection, turning instead to less rigorous sources of data such as the interview and retyping to content validity as their accepted standard for selection practices. The model suggests that no amount of Federal policy regarding selection practice will encourage validation programs within an organization unless the organization is at a stage of maturity sufficient to foster a sophisticated understanding of the potential of valid tests for contributing to a long-range manpower strategy. Organizations are indeed looking for good selection procedures that satisfy their needs for face validity, predictive validity, and fairness, and would adopt such techniques if they were generally available. A major block toward adopting such devices is the perceived need to validate them in each user's organizations in order that they are job-related. Little incentive exists for organization to assimilate the latest and most valid selection practices. Recent advances in selection, documented in academic journals, do not fit the need of organizations which must still take the extra step to insure that a new selection procedure or device will be both valid and fair in its intended application. The relative costs of valdation for an organization are high in terms of both time and resources,

presenting a further obstacle to widespread validation within organizations.

Future policy regarding valid nondiscriminatory selection practice should therefore encourage research by creating incentives for developing new procedures and techniques in conjunction with specialists in selection and competency measurement. Cooperative validation projects could be supported in which organizations would work together with employee selection experts to develop reliable, valid and fair measures of competencies required for successful job performance. Indeed much work is currently being conducted in the insurance and oil industries, funded by a number of separate companies and involving outside experts to carry out just this kind of investigation. The incentives for participating organizations would be minimal assessment development costs, and later unrestricted use of the most workable procedures. The benefits to the funding agencies would be dissemination of high quality research practices to a larger set of organizations and more movement toward using job related standards for fair and equal treatment of qualified job applicants.

Implications for Educators, Employers, and Policymakers

Implications for Educators

On the basis of the literature on different selection techniques, a comparison of techniques, and our survey of selection practices in the field, we can draw certain implications for

education. For example, employers have grown skeptical of the utility of paper and pencil tests for identifying competent job applicants. While test reliability may be strong, the only moderate validity of tests, the test anxiety of applicants, skepticism about face validity, and corporate anxiety over the need to validate tests have led to a down turn in their use since 1972. The result is that student job applicants are less likely to take a selection test than they are to go through an interview or other selection process which they haven't experienced in school. Two of the most common selection techniques are the interview and the application blank, neither of which taps job-related skills by taking any measurements. Both methods are essentially self-report indicators of skills. An applicant may say he or she knows how to operate a drill press but only a job sample, skill test or check with a former employer can verify that report. An interviewer can assess intelligence and interpersonal skills in the employment interview (with or without using rating scales to indicate his or her judgments), and can also obtain a crude reading on literacy, ability to follow directions, and neatness from the application blank. Seldom, however, are these skills the focus of student preparation for a career.

The preceding discussion leads us to the simple conclusion that the skills which play a critical role in surviving the selection process are largely unlike the skills needed to do the job, nor are they skills in which most students have had any significant preparation. An implication for educators,

therefore, is that students need to prepare for the special demands of the selection process. The job applicant who is best prepared to handle a new job is often ignored in favor of one who is more skillful in responding to the questions of an interviewer. As we have seen, even those jobs requiring only psychomotor skills (contrasted with intellectual or interpersonal skills) invariably require an interview, a technique which is strongest in identifying interpersonal skills. Educational institutions would do well to make their students aware of, or better, to educate their students in, the skills required to survive the job search. Further, as each selection method has the potential to provide some unique information about the applicant, educators should also prepare students to take advantage of each method as a way to convey information about themselves. For example, a neat application blank, a well organized and confident resume, and a relaxed but attentive manner in an interview all convey something extra beyond the written or spoken word. Students who have learned to understand these different methods make a better impression on the person charged with the selection decision.

What else can employers tell educators that will help them better prepare students for the world of work? For one thing, educators could avail themselves of analyses of jobs that students are likely to enter upon graduation. However, we should be aware that task analyses, the most common form of job analysis, can supply only a very limited account of competence. This is to say that the skills required to perform

job tasks satisfactorily are often quite different from the competencies that enable one to perform the whole job well. A nurse, for example, needs to acquire many task-specific technical skills to perform at a level of basic competence, but the competencies that supervisors and patients value, such as personal warmth, keeping cool under pressure, and being able to handle a number of tasks simultaneously and efficiently, are not likely to emerge from a task analysis. The example of the nurse is not different from most other complex jobs in that, although task analyses can be worded to derive readily the minimal competencies (competencies needed for survival), task analyses do not lend themselves well to identifying optimal competencies (competencies related to excellence). Viewed in this light, task analyses may not serve the educator well, and schools which teach only to tasks to be performed in jobs trivialize the notion of competence.

Competency identification through task analysis is also a two-edged sword for employers. We have found that relatively few organizations have conducted adequate job analyses on the basis of which selection strategies are chosen, even though the courts have ruled against selection practices that have not been suggested by such analyses. But it seems that employers have come to an intuitive realization that job analysis, and task analysis in particular, does not provide evidence for the qualities they would like to see in applicants such as integrity, interpersonal skill, intelligence, and initiative, to name just a few ill-defined characteristics. All evidence

shows that employers, through the use of a variety of selection devices and their own decision-making process, are trying to select for optimal competencies as well as minimal ones. Unfortunately, organizations do not think in terms of competencies: They are seldom able to articulate just what these important performer characteristics are so that educators can teach them. And, as we have seen, employers are mediocre, at best, in making selection decisions based on these elusive competencies. In the general absence of information, all that educational institutions can do, at present, is guess at many of the competencies that are implied by job analyses.

The reliance on employers for guidance in curriculum design, however, can be carried to an undesirable extreme. The goals of a liberal education must not be confused with the goals of employment. Imparting job related skills is different from enabling an individual to be a more constructive participant in society. Current practices of employee selection are nevertheless driving educational institutions to provide training more closely related to the demands of the job, even though there is no assurance that the qualities on the basis of which one is selected into an organization are at all redeeming in terms of long-term job or career effectiveness. Although educators could learn from available job analyses, they would do better to focus on generic levels of competency which subsume job-related specifics. If more direct translation between job tasks and a program of instruction were

carried out, the result would be a technical school approach to education which, although a worthy alternative for some purposes, would do little to forward the goal of creating a greater awareness of one's environment and development of skills that help one to adapt to changing situations.

It is our conclusion that, with the exception of professional and vocational schools, employers are generally better prepared to impart specific job-related skills, and educational institutions are better prepared to address themselves to generic competencies. Indeed, employers, recognizing the limitations of entry selection practices, are beginning to commit more resources to training and development of people already on the job than to selecting people in advance for the right combinations of skills. Employers cannot realistically expect new employees to come to them fully prepared in terms of the skills required for their first jobs. Within many organizations there exists a current trend away from selection and toward training and career development as a way of acquiring individuals who possess the important competencies for carrying out the work of the organization. This is certainly a costly method of gaining individuals with the appropriate mix of skills, knowledges and abilities, but it may be more efficient in the long run than expecting those combinations of competencies to be available in a job applicant regardless of previous educational experience. Therefore, it appears that educational institutions will be asked more and more to emphasize the development of generic, optimal

competencies. Although educators should be aware of the requirements of the world of work, they would do themselves and their students an ultimate disservice by looking to selection criteria as the only standards against which student performance is to be measured.

Implications for Employers

Current trends in the use of selection techniques, the type of data that these techniques are likely to yield, the reticence of organizations to carry out predictive validation studies and the increased emphasis on training in the job suggest several implications for employing organizations. Employers clearly need to begin undertaking job analyses before implementing any selection screening process. They should begin systematic, empirical documentation of what it takes to do jobs well, rather than the prevalent armchair-theoretic approach to job analysis. Then, documented or undocumented, employers should have a better chance of choosing selection techniques which are appropriate to gathering information on desired job qualities. This is a first step to ascertaining whether hiring decisions made through the use of a given selection program are related to performance in critical job functions.

Employers also need to become more aware of the systematic biases that are introduced into the decision making process by using selection devices that measure factors that are inappropriate and unrelated to job requirements. Employers as a rule

tend to overselect for interpersonal skills and general intelligence, characteristics which have great face validity but may not have any positive relationship to job performance. General intelligence, specifically, has been called into question by McClelland (1973), who noted that this construct appears not to correlate with job performance within career areas, including those professions that appear to demand a high degree of intellectual competence. Systematic job analyses will not, in themselves, insure that these biases will be overcome and given the residual unreliability of even the best selection devices administered under controlled conditions, it is unlikely that these biases will be completely eradicated.

Employers should also recognize that new employees will have many opportunities to train the workers for specific positions once they are on the job. The task of selection, then, becomes more one of identifying needed competencies that are unlikely to be developed on the job rather than trying to account for all competencies needed to do the job at entry. A possible outcome of this strategy is providing greater access to jobs by nontraditional applicants, including women and minorities. Under the state of the practice, previous experience in similar jobs is a powerful factor that determines whether a person will be hired; to cite an example, sales is one area where previous experience weighs in favor of the applicant, regardless of the competence demonstrated by the applicant in previous employment experience. Some of the

minimal competencies required (e.g., product knowledge and knowledge of sales procedures) can be learned ideally on the job, while some of the more critical optimal competencies (e.g., achievement orientation and influence skills) are harder to develop. Employers under most current selection systems over-select for the more easily acquired minimal competencies at entry, placing at a disadvantage women, minorities, and others who have been denied equal access to such jobs in the past.

In theory, there is less likelihood that previously excluded applicant populations possess many optimal competencies to a lesser extent than do the currently favored population. Supporting evidence comes from the experience of the first author in implementing a competency-based sales selection system in a Fortune 500 company. Under the new system, which focused exclusively on selecting for optimal competencies, more women and fewer people with prior sales experience were accepted than was the case under the old system; nevertheless the sales generated by the new applicant group were signficantly greater compared to previous groups, with less than half the employee turnover rate. Not only has the competency-based selection strategy provided greater benefit to the employer, but it has also recognized the distinction between competencies that are developed on the job and those that are needed at entry, while providing greater opportunities for members of applicant subgroups that have been denied access to jobs in the past.

A clear recommendation is that employers and vocational organizations seek to identify the competencies required for the performance of job tasks and then to choose or develop appropriate selection techniques. The competency identification process is not the same as a job analysis. A job yields a listing of general tasks to be performed but does not, in itself, reveal anything about the knowledges, abilities, or other characteristics of the person who performs the tasks well. Rather than to repeat the mistakes of the majority of employers who take only a face validity-based approach to the inference of skills from the job tasks, the most logical approach is to identify, first, those results that a are taken as evidence of satisfactory or outstanding job performance, and then to discover the competencies that are possessed to a greater degree by the satisfactory or outstanding performer than by tne less-than-satisfactory job incumbent. Competency "models" of the good performer in the job would supplant job description as the driving force in employment selection. Specifying the most critical job tasks that are performed, and the competencies needed to perform the whole job well, would go a long way toward improving the utility and validity of selection systems. Additionally, this information would be of enormous benefit to educators who find current job classifications and task analyses largely useless in improving a curriculum.

The long-term benefit of competency-based selection to employers extend beyond ensuring that new hires will be able to

perform their jobs adequately. Properly designed, such selection systems would identify training needs for new employees in addition to applicants' suitability for hire. A key shortcoming of selection systems that rely on biograpical data and personality testing is that it is not clear what the individual can do to develop the needed characteristics, since these sources of data rely heavily on things the person has reported doing in the past. By contrast, the direct measurement of key competencies tells the employer in a straightforward way if the applicant displays the characteristics associated with success and where the skill gaps are that can be developed. Therefore, this new approach to selection is more likely to indicate training needs rather than fixed characteristics.

Implications for Policymakers

The public endorsement of equal employment opoortunity has been received and implemented in ways which imply two different public policy goals. The first and most common interpreation of this endorsement is to see public policy as encouraging and enforcing employment parity, or employing equal or proportional numbers of majority and minority group members. The second interpretation of EEO policy is to see it as encouraging productivity by selecting employees with job-related skills regardless of workers' group status. Both of these policy positions are implied by the often conflicting messages of the courts and enforcement agencies. Which is the intended policy? Once the

Federal Government makes a policy choice between employment parity and productivity, different action steps exist for reaching either of these goals.

The interpretation that the primary goal of public policy is to encourage employment parity is not congruent with what Justice Burger identified as the intent of Title VII of the Civil Rights Act. When he wrote the landmark Griggs decision, (Willie S. Griggs, et al, v. Duke Power Co. Supreme Court Case 124, October term, 1970. Opinion delivered March 8, 1971), he said:

> "Nothing in the Act precludes the use of testing or measuring procedures; obviously they are useful. What Congress has forbidden is giving these devices and mechanisms controlling force unless they are demonstrably a reasonable measure of job performance. Congress has not commanded that the less qualified be preferred over the better qualified simply because of minority origins. Far from disparaging job qualifications as such, Congress has made such qualifications the controlling factor, so that race, religion, nationality, and sex become irrelevant. What Congress has commanded is that any tests used must measure the person for the job and not the person in the abstract."

Unfortunately, and not surprisingly, the public, including many employers, have come to think of equal employment opportunity in terms of rigid affirmative action goals, consent decrees with minority quotas, and the four-fifths (4/5) rule for identifying job discrimination (as described in the 1978 Uniform Guidelines). Employers suspect that compliance officers are vigilantes eager to punish businesses that don't "have their numbers up." They therefore protect themselves by

hiring members of protected classes (i.e., individuals identifiable on the basis of race, creed, color, sex, national origin, age, handicap, marital status). This has often resulted in resentment by employers who feel that they have been pressured to hire individuals on the bases unrelated to job skills. In essence, this trend results in selection that is not performance based. Resources are devoted to documenting demographics of the work force. Thus the current practice of using federal compliance officers for monitoring of organizations does not contribute to improved job-related selection (as identified by Burger), but rather to employment parity which could as easily be accomplished by an unbiased lottery! If simplistic employment parity is the government's policy goal, then current mechanisms should continue to operate unchanged.

On the other hand, if the intent of Congress and the public is to encourage job-related and unbiased selection, then the policy goal really is one of productivity. Job-related selection is related to productivity in the sense that improved performance-based selection can upgrade skills of a work force, reduce needed training resources, decrease attrition by workers who don't fit the job, and reduce the expense of replacing incompetent workers. Prescreening for competence to do a job is cost effective in terms of the reduction it effects on those later, online expenses. These kinds of improvements or productivity gains are obviously of value to all employers and thus ultimately to the nation. The Commmerce Department,

Government Accounting Office, and a cabinet level committee have all devoted time and resources to the issue of productivity. In a 1979 report to the President by the Council of Economic Advisors, the productivity issue arose as an important concern for government and policy. As described above, fair employment practice laws and validation of competency measures for selection have great potential to contribute to a policy of encouraging productivity, but as currently understood, enforced and practiced, they are presently not contributing much. Current enforcement has had the effect of organizations resorting to employment parity strategies, those organizations that have chosen to implement criterion-related validation being in the distinct minority. Clearly new strategies need to be adopted if public policy is to encourage productivity through fair employment practices and selection validation.

One inexpensive procedure for improving productivity is to elucidate the policy of validating competence measures by preparing and distributing an informational pamphlet to employers. The return on investment for putting dollars into validation of selection measures is real. Examples exist of organizations greatly reducing turnover costs with valid selection strategies. For those organizations which are too small in size or unwilling to devote resources needed, cooperative studies and validity generalization studies are low-cost options. While the potential impact of this information sharing is unknown, more employers might be

encouraged to validate selection procedures if validation were phrased not in terms of public policy and social justice, but rather in terms of the employers' self-interest with regard to productivity, return on investment, and the bottom line.

More active efforts could exist to encourage validation of competency measures to use for selection. For example, it is not unreasonable to think of job competency test validation as an investment for which the Internal Revenue Service could give credit. Currently, all personnel system efforts are business expenses, but documentation of applicant flow is not likely to improve an organization's productivity. While efforts to meet the four-fifths (4/5) rule are not investments, money and manpower resources put into design and validation of a selection measure of job competence is an investment in improving the business's productivity. Perhaps some system of allowing investment tax credits for rigorous design and validation of selection strategies could be an incentive for increasing validation (and ultimately productivity).

Small companies without sufficient manpower, expertise or resources for the level of validation effort necessary to earn an investment tax credit could be encouraged to take advantage of other options. For example, organizations with limited resources might be able to qualify for smaller credits provided they invest in and cooperate in a study of similar jobs within one industry, across companies. Cooperative studies for validity generalization is the trend of the future. Organizations with small numbers of jobs can band together with

other organizations, establish the similarity of their jobs, statistically correct for restriction of range and other sample biases, and thereby establish the validity of a procedure to be used for selecting competent applicants. Small businesses who participate in such cooperative efforts might be enabled to do so with the assistance of low-interest SBA loans. This is another possible incentive which could persuade organizations to invest in validation.

As already suggested, using federal compliance officers to monitor organizations with high visibility (to government, public or consumer) is not sufficient to bring about validation of selection systems. Information, appeals to the "bottom line" logic of managers, and financial incentives may facilitate more validation. Another possibility would involve efforts to upgrade the expertise of corporate personnel staff in the area of test design and validation. As one study revealed, the level of corporate expertise in these areas seems to be related directly to the amount of effort organizations put into validation. While organizations without personnel selection or measurement expertise may hire external consultants and finance selection research and development, one way to increase the number or organizations validating might be to build their own staffs. For example, the Federal government could mandate or finance fellowship programs to provide personnel managers or corporate decision makers with an opprortunity to develop some knowledge of or expertise in design, validation and use of

selection measures. On the basis of that knowledge, they could then direct their own organizations in these areas or make a more informed choice to employ qualified external consultants.

When considering policy options and the wisdom of validating, it is important to remember that an organization which chooses to aim for productivity and job-related, validated selection, is also choosing methods which can be used to detect and eliminate adverse impact, as indicated by the earlier discussion of differential validity and test bias. Organizations which are primarily concerned with equitable unbiased selection of previously disadvantaged groups might best serve that goal by vigorous design and validation of competency measures for selection. At the same time they will increase the probability of hiring a competent, productive work force.

V. REFERENCES

Anderson, R. C. The guided interview as an evaluative instrument. Journal of Educational Research, 1954, 48, 203-209.

Anastasia, A. Psychological Testing. New York: Macmillan, 1976.

Arvey, R. D. Unfair discrimination in the employment review: Legal and psychological aspects. Psychological Bulletin, 1979, 86(4), 736-765.

Asher, J. J. The biographical item: Can it be improved? Personnel Psychology, 1972, 25.

Asher, J. J., & Sciarrino, J. A. Realistic work sample tests: A review. Personnel Psychology, 1974, 27(4).

Atkinson, J. W. Toward experimental analysis of human motivation in terms of motives, expectancies, and incentives. In J. W. Atkinson (ed.) Motives in Fantasy, Action and Society. Princeton, N. J.: D. Van Nostrand, 1961.

Bartlett, C. J., & O'Leary, B. S. A differential prediction model to moderate the effects of heterogeneous groups in personnel selection and classification. Personnel Psychology, 1969, 22, 1-17.

Bass, A. R., & Turner, J. N. Ethnic group differences in relationships among criteria of job performance. Journal of Applied Psychology, 1973, 57, 101-109.

Bass, B. M. Situational tests: I. Individual interviews compared with leaderless group discussions. Educational and Psychological Measurement, 1951, 40, 67-75.

Blakeney, R. N., & MacNaughton, J. F. Effects of temporal placement of unfavorable information on decision making during the selection interview. Journal of Applied Psychology, 1971, 55, 138-142.

Boehm, V. B. Negro-white differences in validity of employment and training selection procedures. Journal of Applied Psychology, 1972, 56, 33-39.

Boehm, V. B. Differential prediction: A methodological artifact? Journal of Applied Psychology, 1977, 62, 146-154

Bolster, B. I., & Springbett, B. M. The reaction of interviewers to favorable and unfavorable information. Journal of Applied Psychology, 1961, 45, 97-103.

Bonneau, L. R. An interview for selecting teachers. Dissertation Abstracts, 1957, 17, 537-538.

Borman, W. C. Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. Journal of Applied Psychology, 1975, 60, 556-560.

Boyatzis, R. E. Managing motivation for maximum productivity. APA paper. New York: Sept. 1979.

Boyd, J. B. Interests of engineers related to turnover, selection and management. Journal of Applied Psychology, 1961, 45, 143-149.

Brass, D. J., & Oldham, G. R. Validating an in basket test using an alternative set of leadership scoring dimensions. Journal of Applied Psychology, 1976, 61.

Bray, D. W., & Moses, J. L. Personnel selection. Annual Review of Psychology, 1972, 23, 545-576.

Brown, E. M. Influence of training method and relationship on performance rating. Journal of Applied Psychology, 1968, 52, 195-199.

Campbell, J. T., Dunnette, M. D., Lawler, E. E., III, & Weick, K. E., Jr. Managerial behavior, performance, and effectiveness. New York: McGraw-Hill, 1970.

Campbell, J. T. Tests are valid for minority groups, too. Public Personnel Management, 1973, 2, 70-73.

Campbell, J. T., Otis, J. L., Liske, R. E., & Prien, E. P. Assessments of higher level personnel: II. Validity of the overall assessment process. Personnel Psychology, 1962, 15, 63-74.

Campion, J. E. Work sampling for personnel selection. Journal of Applied Psychology, 1972, 57.

Carlson. Relative influence of a photo vs. factual written information on an interviewer's employment decision. Personnel Psychology, 1969, p. 45.

Carlson, R. E., Thayer, P. W., Mayfield, E. C., & Peterson, D. A. Research on the selection interview. Personnel Journal, 1971, 50, 268-275.

Doppelt, J. E. Progress in the measurement of mental abilities. Education and Psychological Measurement, 1954, 14, 261-264.

Dunnette, M. D. Personnel selection and placement. Belmont, CA: Wadworth, 1966.

Dunnette, M. D. & Borman, W. C. Personnel selection and classification systems. Annual Review of Psychology, 1979, 30, 477-525.

Entwisle, D. R. To dispel fantasies about fantasy-based measures of achievement motivation. Psychological Bulletin, 1972, 77, 377 -391.

Farr, J. L., O'Leary, & Bartlett. Effect of work sample test upon self-selection and turnover of job applicats. Journal o. Applied Psychology, 1973, 58.

Farr, J. L., & York, C. M. Amount of information and primacy-recency effects in recruiting decisions. Personnel Psychology, 1975, 28, 233-238.

Fincher, C. Differential validity and test bias. Personnel Psychology, 1975, 28, 481-500.

Fine, S. A., & Wiley, W. An introduction to functional job analysis. Kalamazoo, MI: Upjohn Institute, 1971.

Ghiselli, E. E. The validity of aptitude tests in personnel selection. Personnel Psychology, 1973, 26, 461-477.

Goldstein, I. L. The application blank: How honest are the responses? Journal of Applied Psychology, 1971, 55(5).

Grant, D. L., & Bray, D. W. Contributions of the interview to assessment of managerial potential. Journal of Applied Psychology, 1969, 53, 24-34.

Guion, R. M. Personnel testing. New York: McGraw-Hill, 1965.

Hakel, M., Dobmeyer, T. W., & Dunnette, M. Relative importance of three content dimensions in overall suitability ratings of job applicants' resumes. Journal of Applied Psychology, 1970, 55.

Hakel, M. D. A legal and psychometric evaluation of selection interviewing practices. Unpublished manuscript, 1977.

Hakel, M. D., Hollmann, T. D., & Dunnette, M. D. Accuracy of interviewers, certified public accountants, and students in identifying the interest of accountants. Journal of Applied Psychology, 1970, 54, 115-119.

Hakel, M., Ohnesorge, J. P., & Dunnette, M. Interviewer evaluations of job applicants resumes as a function of the qualifications of the immediately preceding applicant: An examination of contrast effects. Journal of Applied Psychology, 1970b, 54. 27-30.

Helmreich, R., Bakeman, R., & Radloff, R. Life History QR as a predictor of performance in Navy driver training. Journal of Applied Psychology, 1973, 58.

Heneman, H. G., III. Impact of test information and applicant sex on applicant evaluations in a selection interview. Journal of Applied Psychology, 1977, 62, 524-526.

Heneman, H. G., III, Schwab, D. P., Huett, D. L., & Ford, J. J. Interviewer validity as a function of interviewer structure, biographical data, and interviewee order. Journal of Applied Psychology, 1975, 60, 748-753.

Hakel, M., Dobmeyer, T. W., & Dunnette, M. Relative importance of three content dimensions in overall suitability ratings of job applicants' resumes. Journal of Applied Psychology, 1970, 55.

Hakel, M. D. A legal and psychometric evaluation of selection interviewing practices. Unpublished manuscript, 1977.

Hakel, M. D., Hollmann, T. D., & Dunnette, M. D. Accuracy of interviewers, certified public accountants, and students in identifying the interest of accountants. Journal of Applied Psychology, 1970, 54, 115-119.

Hakel, M., Ohnesorge, J. P., & Dunnette, M. Interviewer evaluations of job applicants resumes as a function of the qualifications of the immediately preceding applicant: An examination of contrast effects. Journal of Applied Psychology, 1970 , 54. 27-30.

Helmreich, R., Bakeman, R., & Radloff, R. Life History QR as a predictor of performance in Navy driver training. Journal of Applied Psychology, 1973, 58.

Heneman, H. G., III. Impact of test information and applicant sex on applicant evaluations in a selection interview. Journal of Applied Psychology, 1977, 62, 524-526.

Heneman, H. G., III, Schwab, D. P., Huett, D. L., & Ford, J. J. Interviewer validity as a function of interviewer structure, biographical data, and interviewee order. Journal of Applied Psychology, 1975, 60, 748-753.

Hinrichs, J., et al. Validity of biographical information blank across national boundaries. Personnel Psychology, 1976, 29.

Hollingworth, H. L. Judging human character. New York: Appleton, 1922.

Hollmann, T. D. Employment interviewers' errors in processing positive and negative information. Journal of Applied Psychology, 1972, 56, 130-134.

Holt, R. R. Clinical and statistical prediction: A reformulation and some new data. Journal of Abnormal and Social Psychology, 1958, 56, 1-12.

Hunt, W. A., Herrmann, R. S., & Noble, H. The specificity of the psychiatric interview. Journal of Clinical Psychology, 1957, 13, 49-53.

Hunter, J. E., Schmidt, F. J., & Hunter, R. Differential validity of employment tests by race: A comprehensive review and analysis. Psychological Bulletin, 1979, 86(4), 721-735.

Huse, E. Assessments of higher level personnel: IV. The validity of assessment techniques based on systematically varied information. Personnel Psychology, 1962, 15, 195-205.

James, L. R., et al. Prediction of artistic performance from biographical data. Journal of Applied Psychology, 1974, 59.

Kane, J. & Lawler, E. Methods of peer assessment. Unpublished paper.

Katz, D., & Kahn, R. The social psychology of organizations (2nd edition). New York: Wiley, 1978.

Keating, E., Paterson, D. B., & Stone, C. H. Validity of work histories obtained by interview. Journal of Applied Psychology, 1950, 34.

Kirkpatrick, J. J., Ewen, R. B., Barrett, R. S., & Katzell, R. A. Testing and fair employment. New York: Unio Press, 1968.

Kauft, E. B. Vocational interests and managerial success. Journal of Applied Psychology, 1951, 35, 160-163.

Kruger, B. R., Shikiar, R. Sexual discrimination in the case of letters of recommendation: A case of reverse discrimination. Journal of Applied Psychology, 1978, 63, 309-314.

Landy, F. J., & Bates, F. Another look at contrast effects in the employment interview. Journal of Applied Psychology, 1973, 58, 141-144.

Langdale, J. A., & Weitz, J. Estimating the influence of job information on interviewer agreement. Journal of Applied Psychology, 1973, 57, 23-27.

Langmuir, C. R., & Kendall, W. E. A logical machine for measuring proolem solving ability. Paper read at XIV International Congress of Applied Psychologists, Copenhagen, 1961.

Latham, G. P., Wexley, K. N., & Pursell, E. D. Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 1975, 60, 550-555.

Levine, J., & Butler, J. Lecture vs. group discussion in changing behavior. Journal of Applied Psychology, 1952, 36, 29-33.

Loevinger, J. Theory and technique of assessment. Annual Review of Psychology, 1959, 10, 287-316.

Maas, J. B. Patterned scaled expectation interview: Reliability studies on a new technique. Journal of Applied Psychology, 1965, 49, 431-433.

Mayer, S. E., & Bell, A. I. Sexism in ratings of personality traits. Personnel Psychology, 1975, 28, 239-249.

Mayfield, E. C. The selection interview--A re-evaluation of published research. Personnel Psychology, 1964, 17, 239-260.

McClelland, D. C. Is personality consistent? In A. T. Rabin (ed.) Symposium in Honor of Henry A. Murray. New York: Wiley, in press.

McClelland, D. C. Testing for competence rather than for "intelligence." American Psychologist, January 1973, 28(1).

McClelland, D. C., et al. The Achievement Motive. New York: Appleton Century-Crofts, 1953.

McClelland, D. C., The Achieving Society. New York: D. Van Nostrand, 1961.

McClelland, D. C., Longitudinal trends in the relation of thought to action. Journal of Consulting Psychology. 1966, 30, 479-483.

McClelland, D. C. & Winter, D. C. Motivating Economic Achievment. New York: Free Press, 1969 & 1971.

McClelland, D. C. Power: the Inner Experience. New York: Irvington, 1975.

McNamara, W. J., & Hughes, J. L. A review of the research on the selection of computer programmers. Personnel Psychology, 1961, 14, 39-51.

Miner, J. B. Executive and personnel interviews as predictors of consulting success. Personnel Psychology, 1970, 23, 521-538.

Miron, D. & McClelland, D. C. The impact of achievement motivation training on small business. California Management Review, 1977, Vol. XXI, No. 4, 13-28.

Mosel, J. N., & Cozen, L. W. The accuracy of application blank work histories. Journal of Applied Psychology, 1953, 36.

Mosel, J. S., & Gohen, H. W. Validity of the employment recommendation questionnaire in personnel selection: Skilled trades. Personnel Psychology, 1958, 11.

Mosel, J. S., & Gohen, H. W. The employment recommendation questionnaire: III. Validity of different types of references. Personnel Psychology, 1959, 12.

Nevo, B. Using biographical information to predict success of men and women in Army. Journal of Applied Psychology, 1976, 61.

Otis, J. L., Campbell, J. H., & Prien, E. P. Assessment of higher level personnel: VII. The nature of assessment. Personnel Psychology, 1962, 15, 441-446.

Owens, W. A., Schumacher, C. F., & Clark, J. B. The measurement of creativity in machine design. Journal of Applied Psychology, 1957, 41, 297-302.

Pace, L. A., & Schoenfeldt, L. F. Legal concerns in use of WgAd applications. Personnel Psychology, 1977, 30(2).

Parry, M. E. Ability of psychologists to estimate validities of personnel tests. Personnel Psychology, 1968, 21, 139-147.

Pashalian, S., & Cressy, W. J. E. The interview: IV. The reliability and validity of the assessment interview as a screening and selection technique in the submarine service. United States Navy Submarine Medical Research Laboratory Report, 1953, 12.

Peters, L. H., & Terborg, J. R. Effects of temporal placement of unfavorable information and of attitude similarity on personnel selection decisions. Organizational Behavior and Human Performance, 1975, 13, 279-293.

Peterson, N. S. & Novick, M. R. An evaluation of some models for test bias. Journal of Education Measrement, 1976, 13, 3-29.

Plag, J. A. Some considerations of the value of the psychiatric screening interview. Journal of Clinical Psychology, 1961, 17, 3-8.

Raines, G. N., & Rohrer, J. H. The operational matrix of psychiatric practice: I. Consistency and variability in interview impressions of different psychiatrists. American Journal of Psychiatry, 1955, 111, 721-733.

Rimland, B. A follow-up analysis of the new composite system for selecting NROTC regular students. United States Navy Bureau of Naval Personnel Technical Bulletin, 1960, No. 60-S.

Roach, D. Double cross validation of WgAd application blank over time. Journal of Applied Psychology, 1971, 56.

Rothschild, W. Putting it all together: A guide to strategic thinking. New York: AMACOM, 1976.

Rundquist, E. A. Development of an interview for selection purposes. In G. A. Kelly (Ed.), New methods in applied psychology. College Park, MD: University of Maryland, 1947.

Sackett, P. R., & Decker, P. J. Detection of deception in the employment context: A review and critical analysis. Personnel Psychology, 1979.

Schmidt, F. L. Berner, J. C. & Hunter, J. E. Racial differences in validity of employment tests: Reality or illusion? Journal of Applied Psychology, 1973, 58, 5-9.

Schmidt, F. L., & Hunter, J. E. Racial and ethnic bias in tests: Divergent implications of two definitions of test bias. American Psychologist, 1974, 29, 1-8.

Schmitt, N. Social and situational determinants of interview decisions: Implications for the employment interview. Personnel Psychology, 1976, 29, 79-101.

Schrader, A. C., & Osburn, H. G. Biographical data taking: Induced subtlety and position of specificity. Personnel Psychology, 1977, 30(3).

Schuh, A. J. Effects of interview rating form content and rater experience on the evaluation of a job applicant. Personnel Psychology, 1973, 26, 251-260.

Schwab, D. P., & Heneman, H. G. Relationship between interview structure and inter-interviewer reliability in an employment situation. Journal of Applied Psychology, 1969, 53, 214-217.

Schwab, D. P., & Oliver, R. L. Predicting tenure with biographical data: Exhuming buried evidence. Personnel Psychology, 1974, 27(1).

Scott, W. D. The scientific selection of salesmen. Advertising and Selling, 1915, 25, 5-6, 94-96.

Scott, W. D., Bingham, W. V. D., & Whipple, G. M. Scientific selection of salesmen. Salesmanship, 1916, 4, 106-108.

Scott, W. D., Clothier, R. C., & Spriegel, W. R. Personnel management (6th Edition). New York: McGraw-Hill, 1961.

Shaffer, D. R., Mays, P. V., & Etheridge, K. Who shall be hired? A biasing effect of the Buckley amendment on employment practice. Journal of Applied Psychology, 1976, 61.

Shaw, E. A. Differential impact of negative stereotyping in employee selection. Personnel Psychology, 1972, 25, 333-338.

Shaw, J. The function of the interview in determining fitness for teacher training. Journal of Educational Research, 1952, 45, 667-681.

Springbett, B. M. Factors affecting the final decision in the employment interview. Canadian Journal of Psychology, 1958, 12, 13-22.

Strong, E. K. Permanence of interest scores over 22 years. Journal of Applied Psychology, 1951, 35, 89-92.

Sydiaha, D. Bales interaction process analysis of personnel selection interviews. Journal of Applied Psychology, 1961, 45, 393-401.

Tenopyr, M. L. Race and socioeconomic status as moderators in predicting machine-shop training success. Paper presented at American Psychological Association, Washington, D.C., 1967.

Thayer, P. W. Something old, something new. Personnel Psychology, 1977, 30(4).

Tiffin, J., & Phelan, R. F. Use of the Kuder Preference Record to predict turnover in an industrial plant. Personnel Psychology, 1953, 6, 195-204.

Trankell, A. The psychologist as an instrument of prediction. Journal of Applied Psychology, 1959, 43, 170-175.

Trites, D. K. Adaptability measures as predictors of performance ratings. Journal of Applied Psychology, 1960, 44, 349-353.

Tucker, D. H , & Rowe, P. M. Relationship between expectancy, causal attribution and final hiring decisions in the employment interview. Journal of Applied Psychology, 1979, 64, 27-34.

Uhrbrock, R. S. Analysis of employment interviews. Personnel Journal, 1933, 12, 98-101.

Ulrich, L., & Trumbo, D. The selection interview since 1949. Psychological Bulletin, 1965, 63, 100-116.

Wagner, R. The employment interview: A critical review. Personnel Psychology, 1949, 2, 17-46.

Ward, L. B. Putting executives to the test. Harvard Business Review, 1960, 38, 6-7, 10, 15, 164-180.

Weitz, J., & Adler. The optimal use of simulation. Journal of Applied Psychology, 1973, 58.

Wexley, K. N., Sanders, R. E., & Yukl, G. A. Training interviewers to eliminate contrast effects in employment interviews. Journal of Applied Psychology, 1973, 57, 233-236.

Wexley, K. N., Yukl, G. A., Kovacs, S. Z., & Sanders, R. E. Importance of contrast effects in employment interviews. Journal of Applied Psychology, 1972, 56, 45-48.

Wiener, Y., & Schneiderman, M. L. Use of job information as a criterion in employment decisions of interviewers. Journal of Applied Psychology, 1974, 59, 699-704.

Wiggins, J. S. Social desirability estimation and "faking good" well. Educational and Psychological Measurement, 1966, 26, 329-341.

Williams, R. L., & Zimmerman, I. M. Accuracy of prediction of military success or failure. United States Armed Forces Medical Journal, 1957, 8, 1478-1494.

Winter, D. G. The power motive. New York: Free Press, 1973.

Winter, D. G. & Stewart, A. J. Power motive reliability as a function of retest instructions. Journal of Consulting & Clincical Psychology, 1977, 45, 436-440.

Winter, D. G., McClelland, D. C., & Stewart, A. J. Competence in College: Evaluating the Liberal University. San Francisco: Jossey-Bass, 1980.(in press)

Winter, D. G. Navy Leadership and Management Competencies: Convergence among Tests, Interviews, and Performance Ratings. Boston: McBer & Company, 1979.

Woodworth, D. G., Barron, F., & MacKinnon, D. W. An analysis of life history interviewer's rating for 100 Air Force captains. United States Air Force Personnel and Training Research Center Technical Note, 1957, No. 57-129.

Wright, O. R., Jr. Summary of research on the selection interview since 1964. Personnel Psychology, 1969, 22, 391-413.

Wright, O. R., Carter, J. L., & Fowler, E. P. A differential analysis of an oral interview program. Public Personnel Review, 1967, 28, 242-246.

Wright, R. A System for Managing Diversity. Cambridge, MA: Arthur D. Little, Inc., 1969.

Zaccaria, M. A., Dailey, J. T., Tupes, E. C., Stafford, A. R., Lawrence, H. G., & Ailsworth, K. A. Development of an interview procedure for USAF officer applicants. United States Air Force Personnel and Training Research Center Development Report, 1956, No. TN-56-43.

"END OF DOCUMENT"